LISIANE ZANELLA

INVESTIGATIONS INTO DEFORESTATION, FOREST FRAGMENTATION AND THE DISTRIBUTION OF THREATENED PLANT SPECIES IN THE SOUTH-EAST OF BRAZIL

LAVRAS-MG

2015

LISIANE ZANELLA

INVESTIGATIONS INTO DEFORESTATION, FOREST FRAGMENTATION AND THE DISTRIBUTION OF THREATENED PLANT SPECIES IN THE SOUTH-EAST OF BRAZIL

Thesis submitted for the degree of Doctor of Philosophy as a Dual PhD with Applied Ecology Postgraduate Program, Federal University of Lavras, Brazil and Lancaster Environment Centre, Lancaster University, United Kingdom.

Supervisors

Dr. Luis Marcelo Tavares de Carvalho

Dr. Alan Blackburn

Dr. Andrew Folkard

LAVRAS-MG

2015

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Zanella, Lisiane.
Investigations into deforestation, forest fragmentation and the distribution of threatened plant species in the South-east of Brazil / Lisiane Zanella. – Lavras: UFLA, 2015.
184 p.: il.

Tese(doutorado)–Universidade Federal de Lavras, 2015.
Orientador(a): Luiz Marcelo Tavares de Carvalho.
Bibliografia.

1. Land use and land cover change. 2. Socio-economic and bio-geophysical factors. 3. Tropical Forests. 4. Priority areas for Conservation. 5. Minas Gerais. I. Universidade Federal de Lavras. II. Título.

LISIANE ZANELLA

INVESTIGATIONS INTO DEFORESTATION, FOREST FRAGMENTATION AND THE DISTRIBUTION OF THREATENED PLANT SPECIES IN THE SOUTH-EAST OF BRAZIL

Thesis submitted for the degree of Doctor of Philosophy as a Dual PhD with Applied Ecology Postgraduate Program, Federal University of Lavras, Brazil and Lancaster Environment Centre, Lancaster University, United Kingdom.

APROVED on 26 of June, 2015

Dr. Milton Cezar Ribeiro	UNESP Rio Claro
Dr. Jos Barlow	Lancaster University and UFLA
Dr. Júlio Louzada	UFLA
Dr. Eduardo van den Berg	UFLA

Dr. Luis Marcelo Tavares de Carvalho
Dr. Alan Blackburn
Dr. Andrew Folkard
Supervisors

LAVRAS – MG
2015

DECLARATION

I hereby declare that this work has been originally produced by myself for this thesis and it has not been submitted for the award of a higher degree to any other institution. Collaborations with other researchers, as well as publications or submissions for publication are properly acknowledged throughout the document.

Lisiane Zanella

Lisiane Zanella,

Lavras, May 2015

Suggested citation: Zanella, L. (2015). Investigations into deforestation, forest fragmentation and the distribution of threatened plant species in the South-east of Brazil. PhD Thesis. Postgraduate Program in Applied Ecology, Lavras, Brazil. Dual PhD with Lancaster Environment Centre, Lancaster University, Lancaster, United Kingdom.

"There is no comparison
between that which is lost by not succeeding
and that which is lost by not trying."

Francis Bacon

To God,
To the love of my life, Victor Hugo,
To my family,
And my true friends...
... For making life worth living!

AKNOWLEDGEMENTS

All my gratitude is for the Lord, who kindly gave me life. For all the great opportunities I experienced, the wonders and the miracles that constantly fill my life. Thank you Lord!

Thanks to my supervisors Prof. Dr. Luis Marcelo (Passarinho), Alan Blackburn and Andrew Folkard, for contributing to the definition of the subject of this thesis and for providing invaluable scientific feedback that improved the quality of this research.

Thank to Postgraduate programs of Applied Ecology at UFLA and of Science of Tropical Environments at Lancaster University for the opportunity of being part of such innovative partnership and for the support throughout the PhD. To the professors in both universities who invested a lot of energy to make the Dual PhD happen, especially Júlio Louzada and Jos Barlow.

I thank all the professors at the Ecology Department, especially Marcelo and Eduardo for their collaboration in this work during the academic reviews. Thanks to the CAPES for support during the PhD.

To the committee members, Dr. Milton Cesar Ribeiro, Dr. Jos Barlow, Dr. Júlio Louzada and Dr. Eduardo van den Berg for their valuable comments and suggestions to this work.

Thank you to all my dear friends from UFLA, who supported me in the most critical moments, who were there even with the distance, my "supervisor siblings", my lab and graduate colleagues, my celebration mates. All friends and colleagues I met throughout the period I am at UFLA and in Lavras. It is impossible to name many names, but each of you played a very important role in this achievement. Thank youuuuu! You will be always in my heart!

To all my dear Lancastrian friends, in special LEC’s Tropical Research Group, thank you! You made my time in the UK simply wonderful! Each of you played a very important role in my life and in this achievement. Thanks you, guys, for the great moments we spent together! I will always remember you! Come to visit me in Brazil.

Thanks to my parents, Ildo and Zaira. You have always given all the support and encouragement necessary and essential to my academic career. You are constantly in my life and constitute in my safe haven.

Thanks to my sisters, Aline and Bruna, my true friends, thank you for the affection, companionship, and all words of enthusiasm.

I also thank the Oliveira's: Wamilton, Janeide, Juliana and Livia, who I consider my second family, composed of such special people.

Finally, thank to my beloved husband, Victor. You are always by my side, in the good and bad moments. You encouraged me, supported me, and gave me the strength to keep going throughout this journey and more. Now we have another happy ending, one more victory, and I am looking forward to the next one. Thank you for being my perfect half, for making me feel complete. I love you with all my heart!

To all those who directly or indirectly helped in the preparation of this thesis. Thank you very much!

RESUMO

Mudanças no uso e cobertura da terra têm resultado em alterações globais, especialmente em ecossistemas tropicais. Investigar fatores que podem estar relacionados a estes distúrbios é essencial para mitigar danos causados à biodiversidade e aos serviços ecossistêmicos. Isso é especialmente importante no caso de biomas tropicais mega-diversos e extremamente ameaçados, como a Mata Atlântica, considerada *hotspot* de biodiversidade. As relações entre as principais ameaças à Mata Atlântica – desflorestamento e fragmentação florestal – e fatores externos não são atualmente bem compreendidas, e compreendê-las é vital para a conservação de sua área remanescente. Igualmente importante, é identificar áreas prioritárias que necessitam de proteção. Compreender a distribuição potencial de espécies ameaçadas, o que também não é bem compreendido atualmente, é fundamental para identificar tais áreas. Esta tese tem como objetivo investigar relações entre desflorestamento e fragmentação florestal, e fatores socioeconômicos e bio-geofísicos na Mata Atlântica no Estado de Minas Gerais, Brasil, e modelar distribuições potenciais de espécies de plantas ameaçadas, que ocorrem neste bioma, no Cerrado e na Caatinga de Minas Gerais.

No Capítulo 2 utilizei a análise de Random Forest (RF) para identificar relações entre fatores socioeconômicos e bio-geofísicos, e métricas de desflorestamento e fragmentação florestal na Mata Atlântica. No capítulo 3, usei RF para investigar como essas relações mudaram quando analisadas em diferentes escalas espaciais. No capítulo 4, eu usei o MaxEnt para investigar a distribuição potencial de espécies de plantas ameaçadas em Minas Gerais, a fim de recomendar áreas prioritárias para conservação com base na maior adequabilidade ambiental das espécies modeladas. Dois projetos em escala ampla realizados em Minas Gerais – o Zoneamento ecológico-econômico de Minas Gerais, ZEE-MG, e o Inventário Florestal – forneceram mais de 300 fatores, métricas de desflorestamento e fragmentação florestal, e parte dos pontos de ocorrência das espécies de plantas ameaçadas. Compilei demais pontos de ocorrência do site SpeciesLink e do banco de dados NeoTropTree.

RF provou ser uma ferramenta eficaz para elucidar as relações entre fatores e métricas, tanto em comparação com a abordagem clássica, como quando usada em múltiplas escalas. No geral, um conjunto de fatores de diferentes categorias foi identificado como tendo relações relativamente fortes com os padrões de desflorestamento e fragmentação florestal. A densidade de estradas foi o fator mais comumente selecionado nos modelos de RF. Observei grande variação nos modelos de RF em diferentes escalas, em que algumas métricas forneceram níveis mais elevados de explicação em escalas mais finas, enquanto outras em escalas mais amplas. Ainda, eu mostrei que os processos de desflorestamento ainda ocorrem na Mata Atlântica de Minas Gerais, mesmo com legislação específica que o torna ilegal neste bioma. Finalmente, eu identifiquei lacunas significativas na proteção das espécies ameaçadas, principalmente na Mata Atlântica. Juntamente com áreas identificadas nos biomas Cerrado e Caatinga, estes locais devem ser considerados como prioridades para a realização de inventários e, após a confirmação de distribuição das espécies existentes, propor novas reservas naturais.

Palavras-chave: Mudanças no uso e cobertura da terra. Fatores socioeconômicos e bio-geofísicos. Florestas tropicais. *Random Forest*. MaxEnt. Áreas prioritárias para Conservação. Minas Gerais.

ABSTRACT

Land use and cover change are resulting in changes across the globe, and especially in tropical ecosystems. Investigating factors that relate to these disturbances is essential if we are to mitigate the damage they are causing to biodiversity and ecosystem services. This is especially important in the case of megadiverse and threatened tropical biomes, such as the Brazilian Atlantic Forest, considered a biodiversity hotspot. The relationships between the greatest threats to the Atlantic Forest – deforestation and forest fragmentation – and external factors that might be related to them are not currently well understood, and comprehending them is vital for the conservation of the forest's remaining area. Equally important is identification of priority areas that need protection. Understanding the potential distribution of threatened species is a key tool for identifying those areas. This thesis aims to investigate the relationships between deforestation and forest fragmentation, and socio-economic and bio-geophysical factors in the Atlantic Forest in the state of Minas Gerais, Brazil, and to model the potential distributions of red-listed plant species, which occur in this biome, and Cerrado and Caatinga of Minas Gerais.

In Chapter 2 I used random forest analysis (RF) to identify relationships between socio-economic and bio-geophysical factors, and deforestation and forest fragmentation metrics in the Atlantic Forest. In chapter 3, I used RF to investigate how those relationships changed when they were analysed at different spatial scales. In Chapter 4, I used MaxEnt to investigate the potential distribution of red-listed plant species in Minas Gerais, in order to recommend priority areas for conservation based on the highest environmental suitability for the species modelled. Two broader-scale projects carried out in Minas Gerais – the ecologic-economical zoning of Minas Gerais, and the vegetation monitoring system dataset – provided the factors, which number more than 300, deforestation and forest fragmentation metrics, and some of the occurrence points of threatened plant species. I compiled further occurrence points from the SpeciesLink Website, and NeoTropTree database.

RF analysis proved to be an effective tool for elucidating the relationships between factors and metrics, both in comparison to the classical approach and when used at multiple scales. Overall, a set of factors from different categories was identified as having relatively strong relationships with patterns of deforestation and forest fragmentation. Road density was the most commonly selected factor in the RF models. I observed extensive variation in the RF models at different scales, with some metrics providing higher levels of explanation at finer scales, while others at larger scales. Additionally, I have shown that deforestation processes are still occurring in the Atlantic Forest of Minas Gerais, even with specific legislation that makes deforestation in this biome illegal.

Finally, I identified significant gaps in the protection of threatened plant species, mainly in the Atlantic Forest. Together with areas identified in the Cerrado and Caatinga, these must be considered as priorities for conducting inventories and, after confirmation of existing species distributions, proposing new natural reserves.

Keywords: Land use and land cover change. Socio-economic and bio-geophysical factors. Tropical Forests. Random Forest. MaxEnt. Priority areas for Conservation. Minas Gerais.

LIST OF FIGURES

LIST OF TABLES

CHAPTER 4

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1 INTRODUCTION

1.1 The Atlantic Forest, the Cerrado, and the Caatinga: human-dominated tropical biomes

1.1.1 The Atlantic Forest

The Brazilian Atlantic Forest constitutes an extremely heterogeneous and unique biome, composed of a set of forest types, as well as natural grasslands, salt marshes, mangroves and associated ecosystems (Ribeiro et al. 2011). It is also a mega-diverse tropical forest, hosting large numbers of species of plants (20,000) mammals (263), birds (936), reptiles (306) and amphibians (475) (Mittermeier et al. 2005). It also has the largest number of endemic species per unit area of any biome in the world, and a notably high diversity of vertebrates (IUCN 2008). For this reason, the Atlantic Forest is considered one of the largest and most important forests in the world (Malhi et al. 2008).

In the 16th century, the Atlantic Forest covered 1.5 million km^2 along the Brazilian coast, and extended west in small islands in Paraguay and Argentina (Galindo-Leal and Câmara 2003, Ribeiro et al. 2009). However, its area has been greatly reduced since then. There is evidence that, even before the Europeans' arrival in Brazil in the early 16th century, the Atlantic Forest was already subject to some level of anthropogenic disturbance by indigenous populations (Dean 1996). The *Tupi* indigenous group dominated the biome for over 1000 years, practicing nomadic slash-and-burn agriculture, but their cultivation system did not impact the Atlantic Forest significantly, and it was able to re-grow vigorously after the collapse of *Tupi* populations (Dean 1996).

Once the Portuguese settlers arrived in Brazil in 1500, they ushered in a resource exploitation chain. Firstly, they overexploited Brazilwood trees

(*Caesalpinia echinata* – currently severely threatened by extinction) impacting nearly 600,000 ha of forest in the first century of European occupation (Dean 1996; Young 2003). They then provided land concessions, encouraging Brazil's occupation and expansion of sugarcane plantations to consolidate their dominance over the territory (Dean 1996). The third economic cycle was gold mining, followed by agricultural expansion to feed the growing population; another 3 million ha of forests were destroyed in the 18th century (Dean 1996). The historical deforestation of the Atlantic Forest culminated with expansion of coffee plantations in the southeastern region of Brazil, from the mid-19th century to the beginning of the 20th century (Dean 1996), followed by expansion of cattle ranching (Young 2003).

In summary, the biome has suffered the consequences of deforestation and fragmentation over five centuries of intense human occupation. Many anthropogenic activities have contributed to the Atlantic Forest degradation (Dean 1996, Young 2003). Recently, disturbances in the biome have been due to anthropogenic activities such as a transition from previous agricultural activities to soybean plantations and forest crops (Campanili and Schaffer 2010), as well as industrialization and urban development (Ribeiro et al. 2011). Over 3000 urban centres built amongst the Atlantic Forest house more than 110 million Brazilians (Pinto et al. 2014). These centres range in size from small villages with simple socio-economic structures to some of the major conurbations of the world (Pinto et al. 2014), where legal and cultural aspects related to the use and dependence on forests also vary from one region to another (Young 2003). Investigating those activities and associated factors is crucial to mitigating further damage to the Atlantic Forest.

The Atlantic Forest is now so highly fragmented and subject to severe anthropic pressure that it is recognized as one of the top-five global biodiversity hotspots (Laurance 2009).The extent of the remaining area of Atlantic Forest is debated, but all studies agree that the amount of remaining area is critically low (about 7-8% of its extent in the 16^{th} century, according to *SOS Mata Atlântica*and INPE (2008); and about 12% according to Ribeiro et al. (2009) who incorporated patches smaller than 50 ha. Trying to reverse this critical scenario, restoration initiatives are under development (Rodrigues et al. 2009, Pinto et al. 2014) in conjunction with the establishment of specific legislation created in 2000 – the Atlantic Forest Law (Law N. 11.428, Brazil2006), to control exploitation and suppression of Atlantic Forest remnants. However, these initiatives seem not to be sufficient to safeguard and restore this biome. Hence, the Atlantic Forest remains at present asa mosaic of forest patches dispersed in a matrix of environments modified by humans. This makes it a perfect example of a fragmented landscape (Ribeiro et al. 2009), which can serve as a test bed for looking for patterns of land use and land cover changes, especially those related to forest fragmentation.

1.1.2 The Cerrado

The Brazilian savanna vegetation is called Cerrado and is the second largest Brazil's major biome, after the Amazon forest (Ratter et al. 1997, Ribeiro and Walter 1998). The Cerrado domain covers more than 2 million km2 (the same size as Western Europe) along the central Brazilian Plateau, extending marginally to Paraguay and Bolivia (Ratter et al. 1997, Alho 2005). Cerrado is composed of unique vegetation types determined primarily by fire or the distribution of soil types (Coutinho 1982), or by a combination of factors such as climate, soil, availability of water and nutrients, geomorphology and topography, latitude, and grazing impact of human activities (Ribeiro and Walter 1998).

The cerrado is one of the richest terrestrial biomes on Earth (Mittermeier et al. 2005). It has a unique fauna and the largest diversity of all savanna floras in the world (ca. 10,000 species) (Ratter et al. 1997, Mittermeier et al. 2005). Overall biodiversity for the Cerrado biome, including all its physiognomic forms, is estimated at 160,000 species of plants, animals, and fungi. Endemicity of cerrado higher plants has recently been estimated at 4,400 species, representing 1.5% of the world's total vascular plant species (Mittermeier et al. 2005). Endemic vertebrates range from 3% (birds) to 28% (amphibians) of the species recorded (Mittermeier et al. 2005). The Cerrados are also unique in that they serve as corridors for species inhabiting neighbouring biomes such as the Amazonian and Atlantic rainforests (Oliveira and Marquis 2002).

Cerrado is the second Brazilian biome that has suffered most severely changes due to human actions, behind the Atlantic Forest (Brazil 2014). The growing pressure to increase the production of meat and grains for exportation has led to a progressive depletion of natural resources of the Cerrado. In the last three decades, the biome has been heavily degraded due to the expansion of the Brazilian agricultural frontier (Brazil 2014). This large-scale transformation of the cerrado landscapes is endangering its biodiversity with habitat fragmentation and even animal extinction. Likewise the Atlantic Forest, the Cerrado is considered a hotspot of biodiversity (Mittermeier et al. 2005).

1.1.3 The Caatinga

Caatinga is a type of tropical seasonal forest, an exclusive biome in Brazil, which biological heritage is unique and not found in any other region in the world. This biome covers an area of about 844.453 km² (11% of the country's territory), covering part of the nine northeastern states and the north of Minas Gerais (Brazil 2014). Hot semi-arid climates (type "BSh" under the Köppen climate classification) are predominant in the Caatinga. However, this biome has strong climate irregularity, recording the most extreme weather values in the country.

The Caatinga has the strongest insolation, the lowest cloud cover, the highest average temperature (between 25 ° C and 30 ° C), the highest evaporation rates, and the lowest rainfall (around 500 and 700 mm annually), with wide spatial and temporal variability (Reddy 1983, Sampaio 2003). This irregular rainfall system along the years generates periodical severe drought issues (Krol et al. 2001).

The Caatinga is divided into ecoregions and has two main vegetation types associated with the heterogeneity of the relief, the climate and the soil of northeast Brazil: forest and non-forest formations. These vegetation types vary accordingly to the leaf deciduousness from evergreen, semi-deciduous to deciduous. Non-forest formations are represented by the thorny deciduous woody vegetation (caatinga stricto sensu), cerrado enclaves, carrasco and other shrub types (Brazil 2005).

In the Caatinga, the pressure of dry weather selected vegetation with protective structures. Several of the woody species store water in their swollen trunks, e.g., Cavanillesia arborea (Lentz 2000). Other plants have their roots practically on the soil surface, to absorb as much rain as possible, as the root effective depth lies among the most sensitive hydrological parameters in water-scarce environments (Güntner and Bronstert 2004).

For a long time, the Caatinga was considered poor in biodiversity and its richness was underestimated due to the recognized gaps in knowledge of this region (Tabarelli and Vicente 2002, Tabarelli and Vicente 2004, Barbosa et al. 2005). However, the number of studies on Caatinga biodiversity has increased considerably over the past decades, and currently, it is known that besides its extreme climate conditions, the Caatinga hosts an impressive faunal and floristic biodiversity and has a high degree of endemism (Albuquerque et al. 2012).

The Caatinga is one of the most often neglected biomes and was for a long time forgotten. Only recently there was a greater concern about the serious situation of biome, because the need for conservation its natural systems, and the wide lack of scientific knowledge (Veloso at al. 2001). Despite its importance, this biome has been extensively modified by human action, and has undergone an intense process of degradation, consequent to the expansion of extensive agriculture and livestock (Tabarelli et al. 2000). Today, less than 2% of the area of Caatinga is protected fully protected conservation units (Tabarelli et al. 2000).

1.2 Land use and cover change: deforestation and forest fragmentation

Land use and cover change (LUCC) is the study of land surface change (Turner II et al. 1990, Lambin et al. 1999). Land cover describes the biophysical attributes of the Earth's surface, while land use defines the human purpose or intent applied to these attributes (Turner II et al. 1990, Lambin et al. 1999). Examples of land cover are forest or desert (Lambin et al. 1999). Agriculture, pasture, or plantations are examples of land use (Geist and Lambin 2002).Land use and cover changes (LUCC) resulting from human activities have transformed the Earth's landscapes for many years, intensifying rapidly over the last three centuries, and accelerating in particular over the last three decades (Lambin and Geist 2006). Tropical deforestation is one of the primary causes of global environmental change (Geist and Lambin 2002). Deforestation occurs when the entire plant biota of an area is removed (Fahrig 2003). It brings negative effects such as biodiversity loss, reduction of genetic potential, scarcity of timber and firewood, climate change, reduction of soil fertility, increased soil erosion, changes in the water regime (Fearnside 2005) and exotic species invasions (Puig 2009).

A major consequence of deforestation is forest fragmentation (Fahrig 2003). Fragmentation refers to the degree of disruption of an originally continuous landscape unit (Metzger 2004) and changes in the habitat configuration as a result of this subdivision and isolation (Fahrig 2003). According to Fahrig (2003), the definition of habitat fragmentation implies four effects of the fragmentation process: (i) reduction in habitat amount (ii) increase in number of habitat patches, (iii) decrease in sizes of habitat patches and (iv) increase in isolation of patches.

When the process of anthropogenic habitat fragmentation occurs (caused by the breakdown of a landscape unit, for example), the landscape structure is modified, resulting in changes in community composition and diversity (Metzger 1999). Habitat loss results in reduction of its total area without it necessarily being broken down. On the other hand, fragmentation leads to the decrease in size, and increases in the number and isolation of remaining patches, but not to the removal of large amounts of habitat (Fahrig 2003). In real landscapes, such as the Atlantic Forest, these processes are often correlated (Fahrig 2003). They change the spatial structure of the landscape and directly affect the persistence of many species (Saunders et al. 1991).

1.3 Landscape ecology: landscape structure and metrics

Landscape ecology can contribute to the understanding of fragmented landscapes (Wu 2006). Landscape ecology is the area within ecology that emphasizes the importance of spatial context for ecological processes and the importance of spatial relations in terms of conservation biology (Metzger 2001).The focus of this approach concerns the effects of landscape spatial structure on ecological processes (Turner 1989).

According to the principles of landscape ecology and the patch-corridor-matrix model, a fragmented landscape presents some basic components such as the matrix, patches and corridors or connecting elements (Forman 1995). The matrix is the most abundant element in the landscape (Forman and Godron 1986, Forman 1995), and is the landscape unit that controls landscape dynamics (Forman 1995). In the Atlantic Forest, the main matrices are extensive areas of pastures, monocultures and urban areas (*SOS Mata Atlântica* and INPE 2008).Patches are homogeneous areas (at a certain scale) of a landscape unit, which are distinguished from neighbouring units and have reduced spatial extent (Metzger 2001). In this study, the habitat patches of interest are the forest remnants left after human colonization. Corridors are homogeneous areas (at a certain scale) of a landscape unit, which are distinguished from neighbouring units and have a linear spatial structure (Metzger 2001). Habitat corridors and stepping-stones are elements that maintain connectivity in a fragmented landscape. Vegetation corridors are strips of vegetation that connect isolated remnants (Bennett 2003) and stepping-stones are isolated, small groups of trees scattered across the matrix (Boscolo et al. 2008). These structures are essential for controlling biological flows in landscapes because they reduce the risk of local extinction and promote new colonization (Metzger 2004). Despite having some potentially negative effects (Hobbs 1992, Lidicker 1999, Anderson and Jenkins 2006), there is no evidence that corridors are consistently detrimental in a way that overcomes their established benefits (Haddah et al. 2014).

Studies of landscape ecology often employ remote sensing data (Turner 2005) that are used to derive land cover maps and to monitor land cover changes (Purkis and Klemas 2011). Landscape structure patterns are analysed by applying metrics (Tischendorf 2001, Turner 1989).These metrics can be quantified for both individual patches and classes, or for the whole landscape (McGarigal and Marks 1995). They are important and useful tools for describing and comparing landscape spatial patterns (Rutledge 2003). Metrics provide valuable information for different applications, serving as tools in environmental monitoring programs (Ståhl et al. 2011); acting as the quantitative linkage between landscape pattern and ecological processes (Dramstad 2009); and allowing for comparisons of different landscapes or studies (Ji et al. 2006). They also are key tools for measuring landscape composition or configuration, and analysing fragmentation and connectivity/isolation of landscape units (Hargis et al. 1998).

Despite being widely used, landscape metrics have some limitations. They are often sensitive to the thematic resolution (i.e. the number of land cover types identified; Bailey et al. 2007), the spatial resolution (i.e., pixel size and extent) (Turner et al. 2001) and classification errors (Hoechstetter et al. 2008). Further, a single metric cannot capture all aspects of landscape patterns (Turner 2005), and one metric may have similar numerical values for different patterns (Tischendorf 2001).

1.4 Deforestation and forest fragmentation related factors

Deforestation remains one of the primary causes of global environmental change (Lambin and Geist 2006), but the question of what drives deforestation remains incompletely answered (NCR 1999). There is evidence from empirical case studies that there is no universal relationship between cause and effect in this respect (Geist and Lambin 2002). Instead, these studies show that deforestation is determined by different combinations of various proximate causes and underlying driving forces in varying geographical and historical contexts (Geist and Lambin 2002), where some of the combinations are robust geographically (e.g. the development of market economies), whereas most of them are region-specific. The same can be expected for forest fragmentation, as it is a consequence of deforestation (Fahrig 2003)

A detailed understanding of the complex set of proximate causes and underlying driving forces affecting forest cover changes in a given location is required prior to any policy intervention (Geist and Lambin 2002). Geist and Lambin (2002) defined underlying causes as comprising regional patterns of economic factors, institutions, national policies, and remote influences, which in turn, drive agricultural expansion, wood extraction, and infrastructure expansion (Geist and Lambin 2002).

They identified four broad clusters of proximate causes: agricultural expansion, wood extraction, infrastructure extension, and other factors (Figure1.1), and five broad clusters of underlying driving forces: demographic, economic, technological, policy and institutional and cultural factors (Figure1.1).In this study, I have accessed information on proximate and underlying causes from a broader-scale project that collected a very large dataset covering a variety of socio-economic and bio-geophysical variables: the ecological-economical zoning of Minas Gerais (ZEE-MG; Scolforo et al. 2008). This project was intended to provide information to government agencies, covering areas of knowledge ranging from poverty indices to fauna vulnerability indices. This dataset has not previously been used to analyse the driving forces of deforestation and forest fragmentation in the Atlantic Forest of Minas Gerais, in spite of the lack of knowledge concerning these matters for guiding conservation and restoration policies.

-

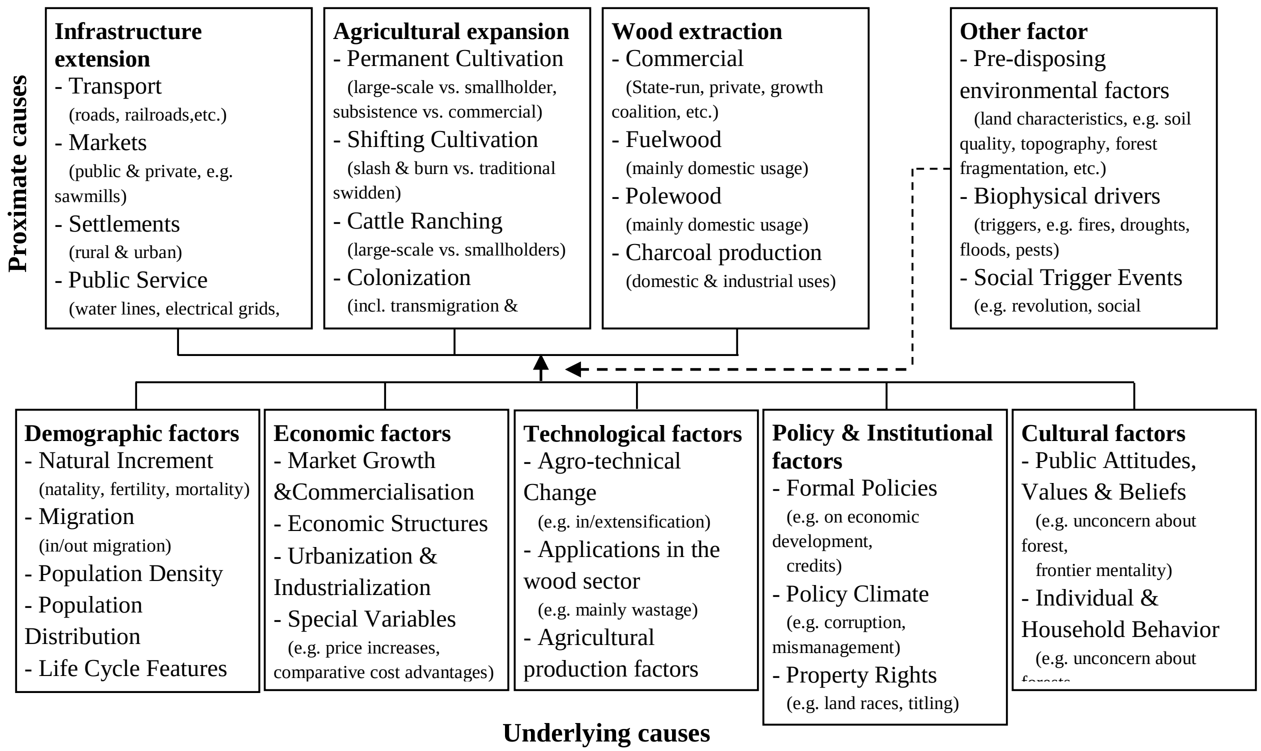


Figure 1.1 Causes of forest decline organized by Geist and Lanbim (2002): five broad clusters of underlying driving forces (or fundamental social processes) underpin the proximate causes of tropical deforestation, which are immediate human actions directly impacting forest cover.

1.5 How to relate deforestation and forest fragmentation to underlying and proximate causes?

The significant progress in the development of statistical modelling tools in recent years has offered a variety of different techniques for investigating factors associated with land use changes. Some studies have used relatively simplistic approaches, such as Mann-Whitney and Kruskal-Wallis tests (Quezada et al. 2013), correlation analyses (Beilin et al. 2014), or statistical redundancy analyses (Parcerisas et al. 2012). Others have applied more robust approaches, combining or comparing different methods. Jaimes et al. (2010) and Gao and Li (2011) compared ordinary least squares regression (OLS) and geographically weighted regression (GWR) to explore factors related to the loss of forest areas and landscape fragmentation. Freitas et al. (2013) combined canonical correspondence analysis (CCA), OLS, GWR and spatial clustering to investigate relationships between land use and land cover changes and environmental and socio-economic variables. Gong et al. (2013) used stepwise multiple regression to examine relationships between urban forest fragmentation metrics and selected socio-economic factors. Bonilla-Moheno et al. (2012) applied a recently developed technique, random forest regression analysis (hereafter, RF; Breiman 2001) to evaluate the effect of environmental, socioeconomic, and demographic variables on woody vegetation trends in Mexican municipalities. However, they did not explore the performance of RF models in this study.

Despite a significant improvement in our understanding of the impacts of LUCC on tropical environments recently (Malhi et al. 2014), there is still no optimal tool for understanding relationships between deforestation/forest fragmentation and socio-economic or bio-geophysical factors. RF analysis has great potential in this respect, since it is capable of identifying complex interactive and non-linear response-predictor relationships, and has excellent predictive performance (Prasad et al. 2006, Smith et al. 2011). Thus, application of RF analysis to disentangle these sorts of relationships may be particularly useful. Furthermore, because many statistical approaches require that variables are normally distributed, most of the research in this field to date has considered only a limited number of independent variables as their starting point. Therefore, modelling approaches need to be further evaluated in terms of the choice of independent and dependent variables, as well as the selection and interpretation of appropriate statistical methods. There is also a need for further studies that include a large number of factors encompassing, as much as possible, all aspects of the socio-economic and bio-geophysical dimensions.

1.5.1 Random forest analysis (RF)

RF (Breiman 2001) is a data mining method widely used in disciplines like bioinformatics (Cutler and Stevens 2006), which has recently gained popularity in ecology (Prasad et al. 2006, Cutler et al. 2007, Wei et al. 2010, Gilbert and Chakraborty 2011). As the name suggests, it uses an ensemble of decision trees with binary divisions, each capable of producing a response when presented with a set of predictor values (Cutler et al. 2007). The available data sets are fed into these decision trees, after which RF uses a classification or regression tree (also known as "CART", Breiman et al. 1984) approach (Prasad et al. 2006) to combine the predictions from all the trees (Cutler et al. 2007). For regression modelling problems, such as the case we considered in this study, the outcome of the RF process that I am primarily concerned with in this study is a measure of the importance of each independent variable for accurate estimation of each dependent variable.

RF has advantages over other methods of identifying relationships between predictor and response (or independent and dependent) variables, in that it does not assume any data distribution, does not require formal selection of predictors, does not over-fit the data (i.e. avoids having noise contaminate the fitting process) and is robust with respect to outliers and unbalanced data (Cutler and Stevens 2006, Prasad et al. 2006). RF is able to analyse large numbers of potential independent variables, and incorporates a method for calculating the importance of each independent variable in the model it provides (Smith et al. 2011). Its main limitations are that its intrinsic flexibility make it somewhat of a "black box" approach (Prasad et al. 2006), but many parameters can be adjusted when performing a modeling. Additionally, it can be very demanding in terms of computational time and resource requirements (Prasad et al. 2006). Nevertheless, recent computational developments have dealt with this limitation.

1.6 Studies in the Brazilian Atlantic Forest: a contextualization

A few studies have attempted to investigate drivers and associated factors of land use and cover changes in the Brazilian Atlantic Forest. Silva et al. (2007) conducted a local scale study and found an indirect influence of topographic relief on forest cover. Teixeira et al. (2009) showed that proximate causes influence the dynamics of deforestation and forest re-growth. They identified that losses in young secondary vegetation and forest were far from rivers, on gentle slopes and near urban areas, while higher forest re-growth rates were near rivers, on steep slopes and far from dirt roads. Freitas et al. (2010) analysed the effects of roads, topography, and land use on forest cover dynamics and demonstrated that forest dynamics were directly related to past road density, past land use (buildings and agriculture expansion), and slope variation. Lira et al. (2012) described LUCC in three Atlantic Forest fragmented landscapes (in São Paulo state) over time and found that LUCC deviated from a random trajectory. Their results also suggested a forest transition in some Atlantic Forest regions. Freitas et al. (2013) used a combination of statistical approaches – multivariate data analysis (CCA), linear regression models (OLS), local spatial regression models (GWR) and spatial clustering procedures (SKATER) – to investigate relationships between LUCC processes and environmental and socio-economic variables in an Atlantic Forest region with an area of ~12,000 km^2 in the state of Rio Grande do Sul. Their findings revealed a competitive and inter-related set of LUCC processes, due to the landscape complexity. More recently, Ferreira et al. (2015) investigated how forest cover and agricultural land use varied in an area of Atlantic Forest in São Paulo state, emphasizing sugarcane expansion.

1.7 Conservation strategies – Protected Areas (PAs)

Of equal importance with studying LUCC effects upon natural environments and the biodiversity and ecosystem services associated, is investing in conservation strategies, which are essential for the maintenance of these environments and the survival of species. Conservation issues are often related to the establishment of PAs, which are considered to be a key strategy of conservation (Butchart et al. 2010). PA design depends on systematic planning, including several steps (Margules and Pressey 2000). Among these, we highlight the identification of threatened species (e.g: Biodiversitas 2005, MMA 2008, IUCN 2014), the identification of hotspots - mega-diversity regions (in terms of species richness, endemism or genetics) which are under greatest threat (e.g. Mittermeier et al. 2005, Mittermeier et al. 2011, Zachos and Habel 2011), and the investigations on the geographical distribution of biodiversity and their current conservation status (Rey Benayas and de la Montana 2003).

Setting aside areas for the preservation of natural values is an old and generalized human practice (Margules and Pressey 2000). These areas are increasingly being established, principally for the protection of biodiversity, including ecosystems, biological assemblages, species and populations (Global Biodiversity Strategy 1992). Although the extent of PA coverage has increased as a response to the biodiversity crisis, the rate of biodiversity loss does not appear to be reducing (Butchart et al. 2010). More efforts are needed to expand PA networks and make conservation initiatives more effective. To do so, basic information such as the geographical distribution of endemic, endangered and exotic species is essential (Whittaker et al. 2005).

1.8 Species distribution modelling

Species distribution models (hereafter, SDM) are considered an important technique that guide conservation practices (Guisan and Zimmermann 2000, Guisan and Thuiller 2005), as the actual geographical distribution of many taxa is little known and most species have only a few occurrence records (Peterson 2006). In the last two decades, SDM received a lot of attention as a tool guiding studies in biogeography, evolution and, more recently, in conservation biology and studies of climate change effects (Guisan and Thuiller 2005).

SDM are obtained through statistical tools that generate the potential geographical distribution of a given group or species based on its occurrence points (Phillips et al. 2006). These models relate points of presence of the species to environmental variables (e.g. temperature, rainfall, altitude and soil types) to predict suitable environments where, in theory, a population can exist and remain viable (Anderson et al. 2003, Guisan and Thuiller 2005). Points of presence are georeferenced coordinates indicating the location where a specimen was collected and/or registered (Anderson et al. 2003). In short, those models can predict, at different scales, sites with environmental characteristics appropriate to the occurrence of this group or species (Phillips et al. 2006). Studies using this modelling approach focused initially on species that are widely distributed, due to the relatively large amount of data available for them (Anderson et al. 2002, Peterson et al. 2008). Recently, they have been expanded to rare, endemic and endangered or threatened species, mainly due to the development of information technology and the availability of a variety of spatial data in different resolutions and scientific collection databases over the internet (Guisan et al. 2006, Williams et al. 2009).

Today, there is a large number of algorithms available for modelling species distributions (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Elith et al. 2006, Muñoz et al. 2009). Therefore, it is important to select the most appropriate method for modelling (Jimenez-Valverde and Lobo 2007), having a well-defined goal, because for each goal, there are different algorithms, modelling and validation techniques.

Amongst all of the techniques available, the maximum entropy approach, "MaxEnt" (Phillips et al. 2004, Phillips et al. 2006) has provided strong evidence of performing better than other methods (Elith et al. 2006, Ortega-Huerta and Peterson 2008). It is able to deal with small sample sizes while remaining effective, and is thus particularly applicable to threatened species, which may have only a few recorded points of occurrence in any given study area (Pearson et al. 2007, Elith at al. 2011). MaxEnt is one of the most commonly used methods for inference of species distributions and environmental tolerances from occurrence data (Phillips et al. 2006). MaxEnt is a maximum entropy based machine-learning algorithm that estimates the probability distribution of a species' occurrence based on environmental constraints (Phillips et al. 2006).

1.9 Thesis' structure and objectives

This thesis is divided into five chapters, comprising an introductory chapter, three chapters detailing substantive pieces of research, and a concluding, synthesis chapter. The thesis was submitted for obtaining the degree of Doctor of Philosophy to satisfy the requirements from the Dual PhD scheme between the Universidade Federal de Lavras – UFLA, Brazil and Lancaster University, England. Thesis' structure follows the formatting guidelines from the Graduate Program of Applied Ecology from UFLA. The three central chapters have all been written in the form of research papers and are intended for individual publication in the peer-reviewed literature: chapter 2 is a multi-authored paper in preparation for submission to Ecological Indicators, chapters 3 and 4 will be submitted for review and publication (target journals are Landscape Ecology,and Biological Conservation, respectively). The thesis is therefore made up of interrelated but stand-alone chapters. The overlap in the text content between chapters has been kept to a minimum; however, there is some common material presented, particularly in the introductions to the papers, methods and references. The specific formatting requirements of the journals have resulted in some minor formatting differences between chapters.

The content of each chapter is summarized as follows:

- Chapter 1 – Introduction. The introductory chapter presents a general introduction covering all themes explored in the thesis. The purpose of this chapter is to provide the reader with an understanding of the broad contextual and theoretical issues that frame this thesis, including introductions to the statistical approaches used to analyse deforestation and forest fragmentation and species distributions.

- Chapter 2 – Random forests explain forest contraction. This chapter presents a comparison between random forest analysis and stepwise multiple regression to investigate factors associated with deforestation and forest fragmentation in the Brazilian Atlantic Forest biome within the state of Minas

Gerais, Brazil.The aims were: (i) to apply random forest analysis in order to elucidate the relationships between deforestation/fragmentation metrics and socio-economic/bio-geophysical factors in the Brazilian Atlantic Forest; and (ii) to compare the performance of random forest analysis with that of multiple linear (stepwise) regression approach to generating understanding of these relationships.

***- Chapter 3 – Multi-Scale Random Forest Analysis for Modelling Relationships between Landscape Pattern and Associated Factors*.** In this chapter, random forest analysis was applied at different spatial scales to investigate the relationships between socio-economic/bio-geophysical factors and deforestation/fragmentation metrics. The main aims were: (i) to quantify and compare deforestation and fragmentation metrics at biome, regional and sub-regional scales in the Atlantic Forest using land-cover maps, and to estimate deforestation rates from 2003 to 2011, and (ii) to identify the forces driving forest fragmentation and deforestation at these different spatial scales in the Atlantic Forest of Minas Gerais, Brazil.

- Chapter 4 –Species distribution modelling demonstrates the need for expansion of protected areas in biodiversity hotspots of Minas Gerais, Brazil. In this chapter, I model endangered plant species distribution, and overlap the models produced with forest remnant data and protected area boundaries. The intention was to evaluate the protection status of these species. The specific aims were: (i) to determine the potential distribution of selected endangered plant species in the Atlantic Forest, Cerrado and Caatinga biomes in Minas Gerais, based on measures of environmental suitability; (ii) to combine state-wide potential distribution models with maps of forest remnants and protected area, in order to assess gaps in the protection of the species studied; and (3) to recommend priority areas for conservation based on the identification of environmentally suitable areas for the species considered in this study.

- ***Chapter 5 – Concluding remarks*. ** This final chapter provides a summary and synthesis of the findings from each of the preceding three chapters, and considers their importance for science and conservation policies in human-modified tropical forests regions, as well as highlighting future research needs.

REFERENCES

ALBUQUERQUE UP.et al. 2012. Caatinga revisited: ecology and conservation of an important seasonal Dry Forest. **Scientific World Journal** 205182:1-18.

ALHO CJR. 2005. Desafios para a conservação do Cerrado em face das atuais tendências de uso e ocupação. In: Scariot A, Sousa-Silva JC, Felfili JM. (Eds.). **Cerrado:** ecologia, biodiversidade e conservação. Brasília: Ministério do Meio Ambiente.

ANDERSON A, JENKINS CN. 2006. **Applying Nature's Design**: Corridors as a Strategy for Biodiversity Conservation New York: Columbia University Press.

ANDERSON RP, GÓMEZ-LAVERDE M, PETERSON AT. 2002 Geographical distributions of spiny pocket mice in South America: insights from predictive models.**Glob Ecol Biogeogr** 11:131–141.

ANDERSON RP, LEWC D, PETERSON T. 2003. Evaluating predictive models of species'distributions: criteria for selecting optimal models **Ecological Modelling**.162:211-232.

BAILEY D. et al. 2007. The influence of thematic resolution on metric selection for biodiversity monitoring in agricultural landscapes **Landscape Ecology** 22:461-473.

BARBOSA M. et al. 2005. Estratégias para conservação da biodiversidade e pesquisas futuras no bioma Caatinga. In: Araújo FS, Rodal MJN, Barbosa MRV (Eds.). **Análise das Variações da Biodiversidade do Bioma com Apoio de Sensoriamento Remoto e Sistema de Informações Geográficas para Suporte de Estratégias Regionais de Conservação**. Fortaleza: Ministério do Meio Ambiente.

BEILIN R. et al. 2014. Analysing how drivers of agricultural land abandonment affect biodiversity and cultural landscapes using case studies from Scandinavia, Iberia and Oceania.**Land use policy** 36:60–72.

BENNETTAF. 2003. **Linkages in the landscape**: the role of corridors and connectivity in wildlife conservation Gland, Switzerland and Cambridge: IUCN. 254p.

BIODIVERSISTAS. 2005. **Revisão da Lista da Flora Brasileira Ameaçada de Extinção.** Available online: http://wwwbiodiversitasorgbr/floraBr/defaultasp (Accessed November 2012) Fundação Biodiversistas.

BONILLA-MOHENO M, MITCHELL-AIDE T, CLARK ML. 2012 The influence of socioeconomic, environmental and demographic factors on municipality scale land-cover change in Mexico.**Regional Environmental Change** 12:543–557.

BOSCOLO D.et al. 2008. Importance of interhabitat gaps and stepping-stones for a bird species in the Atlantic Forest, Brazil.**Biotropica** 40:273-276.

BRAZIL. 2005. **Análise das variações da biodiversidade do bioma caatinga: suporte a estratégias regionais de conservação**. Brasília: Ministério do Meio Ambiente.

BRAZIL. 2006. **Law N. 11428, of 22 of December, 2006.** Regulation for use and protection of native vegetation of the Atlantic Forest biome, and other matters Available from: http://wwwplanaltogovbr/ ccivil_03 /_ato2004 - 2006/2006/lei/l11428htm (Accessed January 2015).

BRAZIL. 2014. Ministério do Meio Ambiente. **Biomas**. Disponível em: (http://www.mma.gov.br/biomas). (Accessed July 2015).

BREIMAN L. et al. 1984. **Classification and regression trees**. Belmont Calif.: Wadsworth.

BREIMAN L. 2001. Random forests. **Mach Learn** 45:5–32.

BUTCHART SHM.et al. 2010 Global Biodiversity: indicators of recent declines.**Science**, 328:1164-1168.

CAMPANILI M, SCHAFFER WB. 2010 **Mata Atlântica**: patrimônio nacional dos brasileiros MMA, Brasília.

COUTINHO LM. 1982. Ecological effects of fire in Brazilian Cerrado. In:. Huntley BJ, Walker BH (Eds.). **Ecology of Tropical Savannas**. Berlin: Springer-Verlag.

CUTLER A, STEVENS J. 2006. Random Forests for Microarrays In: Kimmel A, Oliver B (eds). **Methods in Enzymology**. San Diego: Academic Press.

CUTLER DR. et al. 2007. Random forests for classification in ecology.**Ecology** 88:2783–92.

DEAN W. 1996. **With broadax and firebrand: the destruction of the Brazilian Atlantic Forest.**California: University of California Press.

DRAMSTAD WE. 2009. Spatial metrics – useful indicators for society or mainly fun tools for landscape ecologists? Norsk Geografisk Tidsskrift – **Norwegian Journal of Geography** 63:246-254.

ELITH J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data.**Ecography** 29:129-151.

ELITH J. et al. 2011. A statistical explanation of MaxEnt for ecologists.**Diversity and Distributions**, 17:43-57.

FAHRIG L. 2003. Effects of habitat fragmentation on biodiversity. **Annual Review of Ecology, Evolution and Systematics** 34(1):487–515.

FEARNSIDE PM. 2005. Deforestation in Brazilian Amazonia: History, rates and consequences**. Conservation Biology** 19(3): 680-688.

FERREIRA MP, ALVES DS, SHIMABUKURO YE. 2015. Forest dynamics and land-use transitions in the Brazilian Atlantic Forest: the case of sugarcane expansion. **Regional Environmental Change** 15:365-377.

FORMAN RTT, GODRON M. 1986.**Landscape Ecology.**New York: John Wiley ans Sons, 619.

FORMAN RTT. 1995. Some general principles of landscape and regional ecology.**Landscape Ecology** 10(3):133-142.

FREITAS MWD DE, SANTOS JR DOS, ALVES DS. 2013. Land-use and land-cover change processes in the Upper Uruguay Basin: linking environmental and socioeconomic variables.**Landsc Ecol** 28:311-327.

FREITAS SR, HAWBAKER TJ, METZGER JP. 2010. Effects of roads, topography, and land use on forest cover dynamics in the Brazilian Atlantic.**Forest Forest Ecology and Management** 259:410-417.

GALINDO-LEAL C, CAMARA I. 2003. **The Atlantic Forest of South America**: Biodiversity Status, Threats, and Outlook, Center for Applied Biodiversity Science at Conservation International, Island Press, Washington.

GAO J, LI S. 2011. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression. **Appl Geogr** 31:292–302.

GEIST HJ, LAMBIN EF. 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. **BioScience.**52 (2): 143-150.

GILBERT A, CHAKRABORTY J. 2011. Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida.**Soc Sci Res** 40:273–286.

GLOBAL BIODIVERSITY STRATEGY. 1992. **Policy-makers' guide 1992.** World Resources Institute (WRI), The World Conservation Union (IUCN), United Nations Environment Programme (UNEP) in consultation with Food and Agriculture Organization (FAO), United Nations Education, Scientific and Cultural Organization (UNESCO) Baltimore Maryland.

GONG C. et al. 2013. Determining socioeconomic drivers of urban forest fragmentation with historical remote sensing images. **Landsc Urban Plan** 117:57–65.

GUISAN A, BROENNIMANN O, et al. 2006. Using niche-based models to improve the sampling of rare species. **Conservation Biology** 20:501-511.

GUISAN A, THUILLER W. 2005. Predicting species distribution: offering more than simple habitat models.**Ecol Lett** 8:993-1009.

GUISAN A, ZIMMERMANN NE. 2000. Predictive habitat distribution models in ecology. **Ecol Modell** 135:147-186.

GÜNTNER A, BRONSTERT A. 2004. Representation of landscape variability and lateral redistribution processes for large-scale hydrological modelling in semi-arid areas. **Journal of Hydrology** 297:136 - 161.

HADDAD NM. et al. 2014. Potential negative ecological effects of corridors. **Conservation Biology** 28:1178-1187.

HARGIS CD, BISSONETTE JA, DAVID JL. 1998. The behavior of landscape metrics commonly used in the study of habitat fragmentation.**Landscape Ecology** 13, 167-186.

HOBBS RJ. 1992. The role of corridors in conservation: solution or badwagon? **Trends Ecology Evolutions** 7(11):389-392.

HOECHSTETTER S. et al. 2008. Effects of topography and surface roughness in analyses of landscape structure – A proposal to modify the existing set of landscape metrics.**Landscape Online** 3:1-14.

IUCN. 2008. **Guidelines for Using the IUCN Red List Categories and Criteria Version 7.0**. Prepared by the Standards and Petitions Working Group of the IUCN SSC Biodiversity Assessments Sub-Committee in August 2008. Available from: http://jr.iucnredlist.org/documents/RedListGuidelines.pdf (Accessed November 2014).

IUCN. 2014. **The IUCN Red List of Threatened Species Version 20143.** Available from: http://wwwiucnredlistorg (Accessed November 2014).

JAIMES NBP. et al. 2010. Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression. A**ppl Geogr** 30:576-591.

JI W. et al. 2006. Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics.**Computers Environment and Urban Systems** 30:861-879.

JIMENEZ-VALVERDE A, LOBO JM, HORTAL J. 2007. Not as good as they seem: the importance of concepts in species distribution modelling. **Diversity and Distributions** 14:885-890.

KROL MS. et al. 2001. The semiarid integrated model (SDIM), a regional integratedmodel assessing water availability, 48 vulnerability of ecosystems and society inNE-Brazil. **Physics and Chemistry of the Earth**. 26:529-533.

LAMBIN EF. et al. 1999. **Land-use and land-cover change (LUCC):** Implementation strategy IGBP Report Nº 48, IHDP Report Nº 10, Stockholm, Bonn.

LAMBIN EF, GEIST HJ. 2006. **Land-Use and Land-Cover Change**: Local Processes and Global Impacts – The IGBP Ser. Berlin: Springer-Verlag.
LAURANCE WF. 2009. Conserving the hottest of the hotspots.**Biol Conserv** 142, 1137.

LENTZ DL.2000. **Imperfect balance**: landscape transformations in the Precolumbian Americas. New York: Columbia University Press.

LIDICKER WZ. 1999. Responses of mammals to habitat edges: an overview. **Landscape Ecology** 14(4)333-342.

LIRA PK. et al. 2012. Land-use and land-cover change in Atlantic Forest landscapes.**For Ecol Manage** 278:80–89.

MALHI Y. et al. 2014. Tropical Forests in the Anthropocene. A**nnual Review of Environment and Resources** 39:125-159.

MARGULES CR, PRESSEY RL. 2000. Systematic conservation planning. **Nature** 405:243-253.

MCGARIGAL K, MARKS BJ. 1995. **FragStats:** spatial pattern analysis program for quantifying landscape structure Portland: US Forest Service General Technical Report, 351p.

METZGER JP. 1999. Estrutura da paisagem e fragmentação: análise bibliográfica. **Anais da Academia Brasileira de Ciências** 71:445-463.

METZGER JP. 2001. O que é ecologia de paisagens? **Biota Neotropica** 1:1-9.

METZGER JP. 2004. Estrutura da paisagem: o uso adequado de métricas In: Cullen-Junior L, Rudran R, Valladares-Padua C (Eds). **Métodos de estudos em biologia da conservação e manejo da vida silvestre.** Curitiba: UFPR, p 423-453.

MITTERMEIER R. et al. 2011. **Global Biodiversity Conservation:** The Critical Role of Hotspots. Berlin: Springer Berlin Heidelberg.

MITTERMEIER RA. et al. 2005. **Hotspots revisited:**earth's biologically richest and most endangered terrestrial ecoregions. Boston: University of Chicago Press.

MMA. 2008. **Instrução Normativa no 6, de 23 de setembro de 2008** (ed by Ambiente, M d M). Available from:http://www.mma.gov.br/estruturas/179/_arquivos/179_05122008033615.pdf. (Accessed May 2015).

MUÑOZ MES. et al. 2009. **openModeller:** a generic approach to species' potential distribution modelling. Geoinformatica 15(1):111-135.

NCR. 1999. **Board on Sustainable Development, Policy Division, Committee on Global Change Research**. Global Environmental Change: Research Pathways for the Next Decade. Washington (DC): National Academy Press.

OLIVEIRA PS, MARQUIS RJ. 2002. **The Cerrados of Brazil**. Ecology and natural history of a neotropical savanna.New York:Columbia University Press.

ORTEGA-HUERTA MA, PETERSON AT. 2004. Modelling spatial patterns of biodiversity for conservation prioritization in North-eastern Mexico. **Diversity and Distributions** 10:39-54.

PARCERISAS L. et al. 2012. Land use changes, landscape ecology and their socioeconomic driving forces in the Spanish Mediterranean coast (El Maresme County, 1850–2005). **Environ Sci Policy** 23:120–132.

PEARSON RG.et al. 2007. Predicting species' distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar.**Journal of Biogeography** 34:102-117.

PETERSON AT, PAPES M, SOBERÓN J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modelling.**Ecological Modelling** 213:63-72.

PETERSON AT. 2006. Uses and requirements of ecological niche models and related distributional models.**Biodiversity Informatics** 3:59-72.

PHILLIPS SJ, ANDERSON RP, SCHAPIRE RE. 2006. Maximum entropy modeling of species geographic distributions.**Ecological Modelling** 190:231-259.

PHILLIPS SJ, DUDÍK M, SCHAPIRE RE. 2004. A maximum entropy approach to species distribution modeling In: **Proceedings of the 21st International Conference on Machine Learning**. New York: ACM Press.

PINTO SR. et al. 2014. Governing and Delivering a Biome-Wide Restoration Initiative: The Case of Atlantic Forest Restoration Pact in Brazil. **Forests.**5:2212-2229.

PRASAD AM, IVERSON LR, LIAW A. 2006. Newer Tree Classification and Techniques: Forests Random Prediction Bagging for Ecological Regression.**Ecosystems** 9:181–199.

PUIG H. 2009. **A floresta tropical úmida**. São Paulo: UNESP. 496p.

PURKIS S, KLEMAS V. 2011. **Remote sensing and global environmental change.**Wiley-Blackwell, 384p.

QUEZADA ML. et al. 2013. Land cover changes in the Lachuá region, Guatemala: patterns, proximate causes, and underlying driving forces over the last 50 years. **Reg Environ Chang** 14:1139–1149.

RATTER JA, RIBEIRO JF, BRIDGEWATER S. 1997. The Brazilian cerrado vegetation and threats to its biodiversity. **Annals of Botany** 80: 223-230.

REDDY SJ. 1983. Climatic classification: the semi-arid tropics and its environment:a review. **Pesquisa Agropecuária Brasileira.** 18: 823-847.

REY BENAYAS JM, DE LA MONTANA E. 2003. Identifying areas of high-value vertebrate diversity for strengthening conservation.**Biol. Conserv.** 114:357–370.

RIBEIRO JF, WALTER BMT. 1998. **Fitofisionomias do bioma cerrado**. In: Sano SM, Almeida SP (Eds.). Cerrado: ambiente e flora. Planaltina: EMBRAPA – CPAC.

RIBEIRO MC. et al. 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation.**Biological conservation** 142:1141-1153.

RIBEIRO MC.et al. 2011. The Brazilian Atlantic Forest: a shrinking biodiversity hotspot In: Zachos FE, Habel JC (Eds.) **Biodiversity Hotspots**: Distribution and Protection of Conservation Priority Areas London New York: Springer-Verlag Berlin Heidelberg.

RODRIGUES RR. et al. 2009. On the restoration of high diversity forests: 30 years of experience in the Brazilian Atlantic Forest. **Biol. Conserv.**142:1242–1251.

RUTLEDGE D. 2003. Landscape indices as measures of the effects of fragmentation: can pattern reflect process. **Science Internal Series Archive** 98:1-28.

SAMPAIO EVSB. 2003. Caracterização da caatinga e fatores ambientais que afetam a ecologia das plantas lenhosas. In: Sales VC (Ed.). **Ecossistemas brasileiros**: manejo e conservação. Fortaleza: Expressão.

SAUNDERS DA, HOBBS RJ. 1991. **Nature Conservation 2**: The Role of Corridors. Chipping Norton: Surrey Beatty & Sons.

SCOLFORO JR, OLIVEIRA AD DE, CARVALHO LMT. 2008. **Zoneamento ecológico-econômico do estado de minas gerais**: Componente socioeconômico. Lavras: Editora da UFLA.

SILVA WG.et al. 2007. Relief influence on the spatial distribution of the Atlantic Forest cover on the Ibiúna Plateau, SP. **Braz J Biol** 67(3):403-411.

SMITH SJ, ELLIS N, PITCHER CR. 2011. **Conditional variable importance in R package extendedForest.** Available from http://gradientforestr-forger-projectorg/Conditional-importancepdf (Accessed January 2014).

SOS MATA ATLÂNTICA, INPE. 2008 **Atlas dos remanescentes da Mata Atlântica 2000-2005.**São Paulo:Fundação SOS Mata Atlântica.

STÅHL, G. et al. 2011. National Inventory of Landscapes in Sweden (NILS) – Scope, design, and experiences from establishing a multi-scale biodiversity monitoring system. **Environmental Monitoring and Assessment** 173, 579-595.

TABARELLI M, SILVA JMC, SANTOS AMM. 2000. **Análise de representatividade das unidades de conservação de uso direto e indireto no bioma Caatinga.** Documento Temático, Seminário Biodiversidade da Caatinga (www.biodiversitas.org/caatinga) (Accessed July 2015).

TABARELLI M, VICENTE A. 2002. Lacunas de conhecimento sobre as plantas lenhosas da caatinga. In: Sampaio EVSB, Giulliette AM, Virgírio J, Gamarra-Rojas CFL (Eds). **Caatinga**: Vegetação e Flora. Recife: Associação Plantas do Nordeste e Centro Nordestino de Informações sobre Plantas.

TABARELLI M, VICENTE A. 2004. Conhecimento sobre plantas lenhosas da Caatinga: lacunas geográficas e ecológicas. In: Silva JMC, Tabarelli M, Fonseca MT, Lins LV (Eds.). **Biodiversidade da Caatinga**: Áreas e Ações Prioritárias para Conservação. Recife: Ministério do Meio Ambiente e Universidade Federal de Pernambuco.

TEIXEIRA AMG. et al. 2009. Modeling landscape dynamics in an Atlantic Rainforest region: Implications for conservation. **For Ecol Manage** 257:1219–1230.

TISCHENDORF L. 2001. Can landscape indices predict ecological processes consistently? **Landscape Ecology** 16:235-254.

TURNER II BL. et al. 1990. **The Earth as Transformed by Human Action**: Global and Regional Changes in the Biosphere Over the Past 300 Years. Cambridge: Cambridge Univ Press.

TURNER MG, GARDNER RH, O'NEILL RV. 2001. **Landscape Ecology in Theory and Practice**: Pattern and Process. Springer, New York.

TURNER MG. 1989. Landscape ecology: the effect of pattern on process. **Annual Review of Ecology, Evolution and Systematics** 20(1):171-197.

Turner MG. 2005. Landscape ecology in North America: Past, present, and future. **Ecology** 86:1967-1974.

VELOSO HP, RANGEL-FILHO ALR, LIMA JCA. 1991. **Classificação da vegetação brasileira, adaptada a um sistema universal**. Rio de Janeiro:IBGE, Departamento de Recursos Naturais e Estudos Ambientais.

WEI C-L. et al. 2010. Global patterns and predictions of seafloor biomass using random forests. **PLoS One** 5:e15323.

WHITTAKER RJ, et al. 2005. Conservation Biogeography: assessment and prospect. **Diversity and Distributions** 11:3-23.

WILLIAMS JN. et al. 2009. Using species distribution models to predict new occurrences for rare plants.**Diversity and Distribution** 15:565-576.

WU J, HOBBS R. 2002. Key issues and research priorities in landscape ecology: an idiosyncratic synthesis. **Landscape Ecology** 17: 355–365.

YOUNG CEF. 2003. Socioeconomic causes of deforestation in the Atlantic Forest of Brazil. In: Galindo-Leal C, Câmara IG (Eds.). **The Atlantic Forest of South América:** biodiversity status. Threats and Outlook. Washington:Island Press.

ZACHOS FE, HABEL JC. 2011. **Biodiversity Hotspots**: Distribution and Protection of Conservation Priority Areas. Berlin Heidelberg, London. New York: Springer-Verlag.

CHAPTER 2

RANDOM FORESTS EXPLAIN FOREST CONTRACTION

Publication status: *In prep.* for submission to *Ecological Indicators*

ABSTRACT

Context: Anthropogenic transformations of land cover have changed the Earth's surface globally. Mitigating the damage caused by these changes to the natural environment requires understanding of relationships between spatial distributions of land cover and socio-economic and bio-geophysical parameters.
Objective: To assess the value of applying random forest analysis (RF), a recently developed machine-learning technique, to relating deforestation and forest fragmentation to socio-economic and bio-geophysical variables, in the Brazilian Atlantic Forest of Minas Gerais, Brazil.
Method: A vegetation-monitoring project provided land cover maps, from which we derived deforestation and forest fragmentation metrics. An ecologic-economical zoning project provided more than 300 socio-economic and bio-geophysical variables. We used RF to identify relationships between these sets of variables, and compared its performance in this task to that of a more traditional multiple linear regression approach.
Results: Our investigation showed that RF modelled relatively well variance in metrics describing the density and isolation of forest patches, but was relatively poor at modelling variance of overall rates of deforestation and forest patch shapes, where the multiple linear regression performed better. RF also identified variables describing distances of forest patches from elements of natural and man-made infrastructure, accessibility and topography as being most closely associated with patterns of deforestation and forest fragmentation. In contrast, it found variables associated with economic productivity and social institutions to have little relationship with these patterns.
Conclusions: We found that RF was better at explaining variations in metrics describing patch patterns, while it appears to have been less robust at capturing variations in metrics describing broader landscape structure, possibly due to differences in how these metrics vary in space. We conclude that RF provides a promising methodology for elucidating the relationships between socio-economic/ bio-geophysical factors and patterns of deforestation/forest fragmentation.

Keywords: Land use and land cover change. Deforestation. Forest fragmentation. Socio-economic and bio-geophysical factors. Brazilian Atlantic Forest. Tropical forests. Machine-learning technique. Stepwise Multiple Regression. Minas Gerais State.

2.1 Introduction

A large proportion of the Earth's surface has been transformed by anthropogenic land use activities in recent centuries. Land use and land cover change (hereafter, LUCC) was once considered a local environmental issue, but is becoming globally important due to its increasingly widespread effects upon natural environments. Comprehending these effects requires, in part, the understanding of relationships between variations in socio-economic and bio-geophysical factors associated with the LUCC with which they co-occur. However, comprehending these relationships is difficult because LUCC is influenced by multiple factors acting across different scales of space and time (Geist and Lambin 2002). Therefore, it is necessary to design studies of these relationships carefully so that inferences are reliable. Unreliable conclusions can lead to distorted management recommendations, resulting in missed conservation opportunities, and a waste of resources and time.

Several studies have investigated relationships between LUCC and a wide variety of socio-economic and environmental factors, using a range of statistical techniques. Some studies have used relatively simplistic approaches, such as Mann-Whitney and Kruskal-Wallis tests (Quezada et al. 2013), or correlation analyses (Beilin et al. 2014). Others have applied more robust approaches, combining or comparing different methods.Parcerisas et al. (2012) used statistical redundancy analyses (RDA) to analyse drivers of land cover changes, and consequently, changes in both structural and functional landscape properties. Jaimes et al. (2010) compared ordinary least squares regression (OLS) and geographically weighted regression (GWR) to explore factors related to the loss of forest areas in Mexico and concluded that GWR models represent a significant improvement over OLS models for this purpose. The results from Gao and Li (2011) comparing OLS and GWR to investigate the relationships between landscape fragmentation and related factors agreed with this conclusion. Freitas et al. (2013) presented a combination of canonical correspondence analysis (CCA), OLS, GWR and spatial clustering procedures in order to investigate the relationships between LUCC and environmental and socio-economic variables. Gong et al. (2013) used stepwise multiple regression models to examine relationships between urban forest fragmentation metrics and selected socio-economic factors. Most of the aforementioned studies considered a limited number of potential independent variables that have normal distribution, as this is the basic requirement for using the respective parametric approaches. Therefore, modelling approaches must be further evaluated in terms of the choice of independent and dependent variables, as well as the selection and interpretation of appropriate statistical methods. There is also a need for further studies that include a large number of factors encompassing, as much as possible, all aspects of the socio-economic and bio-geophysical context within which LUCC is taking place.

In this study, we investigate the application of a relatively novel statistical technique – the machine learning algorithm known as "random forest analysis" (RF hereafter, Breiman 2001) – to the task of identifying relationships between a large set of socio-economic and bio-geophysical candidate independent variables, and dependent variables which quantify the current patterns of deforestation and forest fragmentation of the Brazilian Atlantic Forest in the state of Minas Gerais, Brazil. To our knowledge, only one previous study (Bonilla-Moheno et al. 2012) has investigated relationships between LUCC and socio-economic factors using RF. They found it to be a promising statistical approach for this type of study, but used it only to investigate patterns of overall distribution of land cover classes. Here, we extend the application of RF to consider its ability to identify relationships with variables that describe patterns of forest fragmentation. Moreover, this study considers an unusually large set of more than 300 socio-economic and geo-biophysical independent variables. In order to assess the performance of RF, we compared its results with stepwise multiple linear regression, a classical statistical approach, to the same datasets.

2.2 Methods

2.2.1 Study area

The state of Minas Gerais is located in South-eastern Brazil between latitudes 14º 03' 28" S and 23º 07' 02" S and longitudes 51º 07' 02" W and 39º 49' 58" W. It covers an area of 58,652,212 ha and is split into 853 municipalities, ranging in area from 285 ha to 1,071,696 ha. It has three biomes within its limits: Cerrado, Caatinga and Atlantic Forest (IBGE 2004). The study area comprises the 518 municipalities which fall entirely within the largest contiguous area of the Atlantic Forest biome within the state, and encompasses 34% (19,904,146 ha) of Minas Gerais (Figure 2.1) (IBGE 2015). This study site is appropriate for the purposes of this project as there is a wide variability across the municipalities in the magnitude of deforestation and fragmentation and in the socio-economic and bio-geophysical variable values. Indeed, Minas Gerais is oneof the few Brazilian states that has estimates of both forest cover change, and most of the socio-economic/bio-geophysical variables available at the municipality scale.

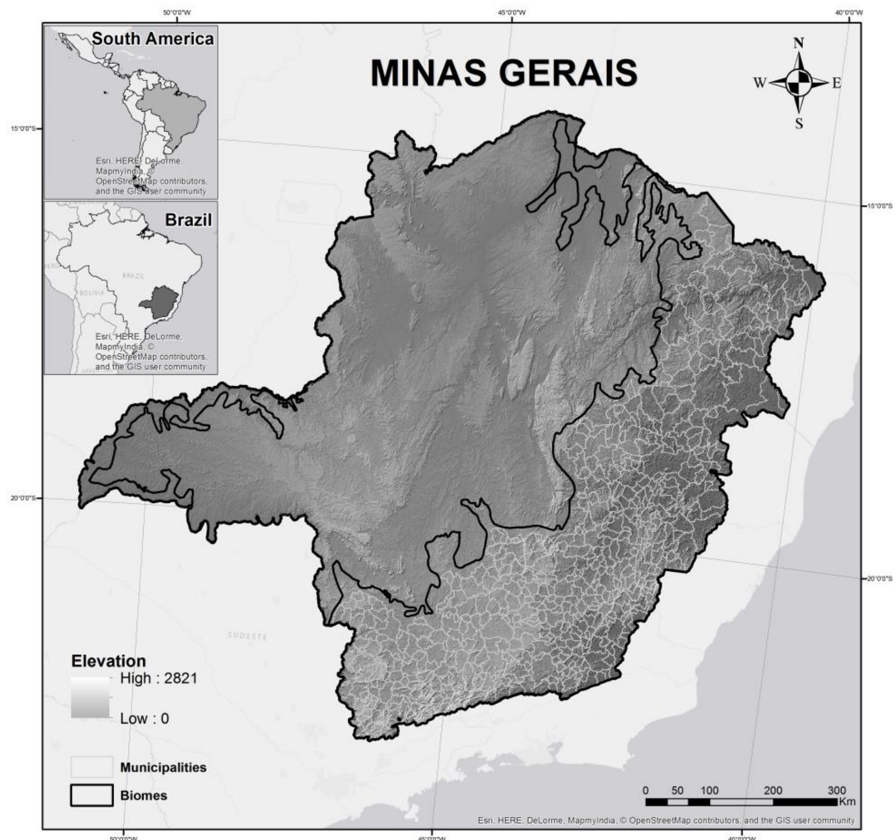


Figure 2.1 Atlantic Forest Biome (Minas Gerais, BR) and the 518 municipalities used in this study. Elevation in metres. The inset maps on the left show the location of Brazil in the South America in the upper map, and the Minas Gerais State within Brazil in the lower map.

2.2.2 Variable selection

This work used large datasets provided by two broader-scale projects carried out in Minas Gerais State, Brazil. The deforestation and forest fragmentation metrics were derived from the vegetation monitoring system dataset (Scolforo and Carvalho 2006, Carvalho and Scolforo 2008,Carvalho and Scolforo - unpublished data), which comprises land cover maps from 2003 to

2011. The socio-economic and bio-geophysical variables were derived from the ecologic-economical zoning of Minas Gerais, ZEE-MG (Scolforo et al. 2008).

A deforestation metric (DEFOR, hectares – Table 2.1) – the total area of land deforested between 2003 and 2011 – was calculated for each municipality using digital change detection applied to Landsat images from the vegetation monitoring system dataset (Scolforo and Carvalho 2006, Carvalho and Scolforo 2008, Carvalho and Scolforo – unpublished data).DEFOR was normalized to the remaining forest area within each municipality.

To quantify forest fragmentation, we used the 2011 land cover map from the vegetation monitoring system dataset (Scolforo and Carvalho 2006, Carvalho and Scolforo - unpublished data). A set of 225 landscape metrics from class and landscape levels from all of the different categories available in FragStats 4.0 (McGarigal et al. 2012) were calculated for each of the 518 municipalities considering the forest cover configuration in 2011. These were then passed through a three-stage filtering process to provide a tractable set of dependent variables for use in our analysis of statistical approaches. Firstly, noting that metrics in datasets such as this can be highly correlated (Riitters et al. 1995), we selected a subset of uncorrelated metrics based on correlation analyses with a confidence threshold of 0.01, discarding those which were strongly correlated with selected variables (and therefore deemed to be redundant) after careful consideration of their ecological meaning. When two or more variables were significantly correlated, the selection criteria to choose one of them were mathematical simplicity and an intuitive judgment of their explanatory power in terms of ecological meaning. Secondly, we chose metrics from the remaining subset that were commonly used in literature found via a search on the Web of Knowledge website (http://wok.mimas.ac.uk/). The search was carried out from 2011 to June 2013, using the key-words "landscape metrics" and/or "landscape indices". This search yielded 48 papers, of which four were found, on inspection, to be out of scope, and we had no access to another five. The papers consulted in the review can be seen in the Appendix Chapter 2 (A2 – List S1). Finally, we verified the normality of the residuals from linear models (see the section *2.2.4 Stepwise multiple linear regression* for more details) and those metrics which had non-normally distributed residuals were discarded to enable comparative analysis of the random forest method with classical, parametric multiple regression, which requires normally distributed variables. The result of this filtering process was selection of three landscape metrics representing fragmentation: the mean Euclidean nearest-neighbour distance (ENN), a measure of patch's isolation from each other; the landscape shape index (LSI), a measure of forest patch shape complexity; and the patch density (PD), a measure of forest spatial structure (Table 2.1).

Table 2.1 Descriptions of deforestation and forest fragmentation metrics (dependent variables).

Metric	**Category**	**Formulae**	**Description (unit)[a]**
Total deforestation (DEFOR)	Deforestation quantification	$\mathrm{DEFOR} = \sum_{j=1}^{n} a_{ij}$	DEFOR equals the sum of the areas of all patches of deforestation between 2003 and 2011, in hectares. a_{ij} = area (m^2) of patch number *j* of cover type *i* (in this case deforested land); n = total number of patches.DEFOR was normalized to the remaining forest area within each municipality.
Mean Euclidean Nearest-Neighbour (ENN)	Forest patch isolation	$\mathrm{ENN} = \frac{\sum_{j=1}^{n} h_{ij}}{n_i}$	ENN equals the mean distance to the nearest neighbouring patch of forest, based on shortest edge-to-edge distance. h_{ij} = distance (m) from patch *j* to nearest neighbouring patch of the same type (*i*, in this case forest). n_i = number of patches of cover type *i* (forest).

"Table 2.1, conclusion."

Metric	Category	Formulae	Description (unit)[a]
Landscape shape index (LSI)	Shape complexity of forest patches	$\mathrm{LSI} = \frac{0.25\sum_{k=1}^{m} e_{ik}^{*}}{\sqrt{A}}$	Landscape shape index (LSI) reflects the shape and complexity of the patches by measuring the perimeter-to-area ratio for the landscape as a whole, which increases with fragmentation. e_{ik}^{*} = total length (m) of edge in landscape between patches of cover types *i* and *k*; *m* = number of different cover types; *A* = total landscape area (m^2).
Patch density (PD)	Forest spatial structure	$\mathrm{PD} = \frac{n_i}{A}(10{,}00000)$	Patch density increases with a greater number of patches within a reference area and therefore reflects landscape fragmentation.

[a] Details can be found in McGarigal et al. (2012).

Socio-economic and bio-geophysical variables were obtained from the ZEE-MG database. The years for which these variables were collected were limited by the availability of information from national agencies, and ranged from 2003 to 2006. Based on data availability, and following Scolforo et al. (2008), socio-economic variables from four categories – production, natural, human and institutional – were used. Variables from further three categories of bio-geophysical factors – topography, distance and accessibility – were also selected. This gave an initial list of more than 300 candidate independent variables. Descriptions of how these variables were calculated can be found in Scolforo et al. (2008). From this list, a tractable sub-set of variables was derived using the first step from the filtering process described above for the forest fragmentation variables. As a result, a total of 32 socio-economic and bio-geophysical variables were selected as independent variables for use in our comparative analysis of statistical approaches (Table 2.2).

Table 2.2 Socio-economic and bio-geophysical variables that will be used as independent variables.

Category	Acronym	Description[a]
Productive component	VA_agri	Value added by agriculture sector adjusted by inflation in 2004 (R$[b])
	GGR_GPM	Geometric growth rate of total gross product per municipality (1999-2003)
	GPM	Total gross product per municipality in 2004 (R$)
Natural component	For_crops*	Amount of forest crops per municipality in hectares
	Perm_crops*	Amount of permanent crops per municipality in hectares
	Annu_crops*	Amount of annual crops per municipality in hectares
	Cov_areas*	Amount of area covered by reservoirs per municipality in hectares
	Prot_areas*	Amount of protected areas per municipality in hectares
	Rural_fam*	Number of rural family farms per municipality. Rural family farms are those in which labour is performed by family members only.
	Min_comp	Index of financial compensation for mineral extraction in 2005 (R$)
	ICMS_eco	Financial compensation given to municipalities that have land use restrictions due to protected areas (Ecological ICMS) in 2005 (R$)
	Min_conc	Number of mines in operation by municipality
Human component	Occu_rate	Intensity of usage of available land for economic use, obtained from the total area of the municipality available for economic activities minus protected and flooded areas.
	Unemp	Unemployment rate per municipality in 2005
	Emp	Geometric growth rate of formal employment per municipality between 2000 and 2005
	Den_pop	Population density per municipality in 2004
	Urb_pop	Percentage of urban population per municipality in 2004
	Income	Per capita income per municipality (R$ / per capita)
Institutional component	Law_enf	Law Enforcement Capacity measures the ratio between the number of citizens and the number of Military and Civil Policemen, Judges, District Attorneys and Public Defenders in the municipality. It ranges from 0 to 1 (0 - no capacity; 1 - best capacity)

"Table 2.2, conclusion."

Category	Acronym	Description[a]
Topography	Alt_mean	Mean altitude per municipality in meters
	Slo_min	Minimum slope per municipality in degrees
	Slo_mean	Mean slope per municipality in degrees
	Slo_mean_d	Mean slope of areas deforested between 2003 and 2011, per municipality in degrees
	Rock	Predominant rock type covered by forests in the municipality
	Soil	Predominant soil type covered by forests in the municipality
Accessibility	Roads_den	Road density per municipality (km/1.000 km^2)
	Rail_den	Railway density per municipality (km/1.000 km^2)
Distance factors	MinDist_sm	Mean distance of forest patches within the municipality to the closest steel mill in kilometres
	MinDist_ri	Mean distance of forest patches within the municipality to the closest river in kilometres
	MinDist_nr	Mean distance of forest patches within the municipality to the closest protected area in kilometres
	MinDist_ro	Mean distance of forest patches within the municipality to the closest road (highway) in kilometres
	MinDist_re	Mean distance of forest patches within the municipality to the closest reservoir in kilometres

[a] The full description on how the variable were calculated and their description can be found in Scolforo et al. (2008).
[b] Brazilian real currency
*Variables scaled to the municipality area prior modelling to allow comparison.

2.2.3 Random forest analysis (RF)

Random forest analysis is a machine-learning technique that may be used for predictive modelling of multiple outputs from large input datasets. It is used widely in bioinformatics (Cutler and Stevens 2006), and has recently gained popularity in ecology (Prasad et al. 2006, Cutler et al. 2007, Wei et al. 2010, Gilbert and Chakraborty 2011). In short, RF, as the name suggests, uses an ensemble of decision trees with binary divisions, each capable of producing an output when presented with a set of input values (Cutler et al. 2007). For regression modelling problems, such as the case we considered in this study, the tree response is an estimate of dependent (output) variable values derived from the given values of a set of independent (input) variables.

RF uses a classification or regression tree approach (also known as "CART"; Breiman et al. 1984), to build a number of decision tree models from randomly selected subsets of training samples and independent variables (Cutler et al. 2007). Model fitness is then examined using validation data that is not in the training sub-sample; hence, cross-validation with external data is not necessary. The outputs from all of the trees are then averaged, which provides predictive accuracy and low bias (Breiman 2001). RF has further advantages in that it does not assume any data distribution, does not require formal selection of predictors, does not over-fit the data (i.e. avoids having noise contaminate the fitting process) and is robust with respect to outliers and unbalanced data (Cutler and Stevens 2006,Prasad et al. 2006). Its main limitations are that its intrinsic flexibility make it somewhat of a "black box" approach, but many parameters can be adjusted when performing a modeling. Additionally, it can be very demanding in terms of computational time and resource requirements (Prasad et al. 2006). Nevertheless, recent computational developments have dealt with this limitation.

We used the R package "extendedForest" provided by the Gradient Forest project (Smith et al. 2011, Ellis et al. 2012) to carry out RF analysis. This package was developed for use in ecological studies of species distributions. It integrates results from RF analyses for a number of individual species distributions into results that enable prediction of multiple species distributions (Smith et al. 2011, Ellis et al. 2012). In addition, it is able to analyse large numbers of potential independent variables, and incorporates a method for calculating the importance of each independent variable in the model it provides (Smith et al. 2011). In our study, we extended the application of the extendedForest package by using the deforestation and forest fragmentation metrics described above (i.e. DEFOR, ENN, LSI, and PD) as our dependent variables, in place of the species distributions used in the application for which it was originally developed.

2.2.4 Stepwise multiple linear regression

From a wide range of possible approaches, we selected stepwise multiple linear regression (hereafter, STEP) as a comparator method against which to assess the performance of RF. This type of technique is arguably the most common approach to data-based prediction and simulation tasks. For situations in which the number of variables is high, as is the case here, it is appropriate to incorporate into the modelling process a method for selecting only those independent variables that contribute most strongly to the predictive model delivered. The STEP approach to multiple regression is a routine technique for achieving this (see, for example, Efroymson 1960, Hocking and Mar 1976, Furundzic 1998, James et al. 2013). Despite having a number of weaknesses, notably bias in parameter estimation, inconsistencies among model selection algorithms, and an inappropriate focus on a single best model (Burnham and Anderson 2002, Kadane and Lazar 2004, Whittingham et al. 2006), it is used widely within ecology and landscape studies (Whittingham et al. 2006).

The stepwise method combines forward selection and backward elimination procedures (Venables and Ripley 2002, James et al. 2013). It proceeded by first setting up an initial model incorporating a subset of the candidate independent variables. Then, this model was iteratively altered by adding significant variables and/or removing insignificant ones, in a process called the stepping procedure. A variable that entered at an early stage may have become superfluous at later stages because of its relationship with other variables subsequently added to the model (Kleinbaum et al. 1998). To check this possibility, at each step a partial F test is carried out for each variable currently in the model, regardless of the stage at which it was entered. The whole process is repeated until no more variables could be added or removed, which means that the model is optimized, or when a specified maximum number of steps is reached. Many statistical methods are available to test the stability and validity of the final regression model. We used the adjusted square of the correlation coefficient (adjusted R^2) and the AIC (Akaike Information Criteria) to assess our final model. The AIC was also used to calculate relative variable importance. Implementation was based on the *dredge* function for automated model selection, which is available as an R package *MuMIn* (Barton 2014). It calculates AIC values for models with all possible combinations of predictor variables and ranks the models based on the calculated values. We determined the relative importance (Burnham and Anderson 2002) of each independent variable selected in the STEP models based on AIC weights (*importance* function in *MuMIn*). The relative importance values were converted to percentages for comparison.

2.3 Results

2.3.1 Random forest analysis

The RF analysis provides evidence of relevant relationships between the independent variables (socio-economic and bio-geophysical factors – Table 2.3; Figure 2.2) and the dependent variables (deforestation and forest fragmentation metrics). However, the outputs also imply that there is restricted explanatory power in the independent variables and reasonable variability in the dependent variables across the municipalities that is not explained by the independent variables considered here.

Table 2.3 Summary of the outputs from the random forest (RF), and stepwise multiple regression (STEP) analyses. Independent variables (shown in the body of the table) are defined in Table 2.2, and dependent variables (shown on the left hand side) are defined in the text.

	RF				STEP			
	Variables	**+/-[a]**	**%imp[b]**	**%var[c]**	**Variables**	**+/-**	**%imp**	**%var**
DEFOR[d]	Min_comp	-	1.35	11.23	MinDist_sm	-	3.40	30.9
	Roads_den	+	1.24		Income	-	3.40	
	MinDist_sm	-	1.23		Prot_areas	+	3.33	
	Slo_mean	-	1.20		MinDist_nr	+	3.26	
	ICMS_eco	+	1.08		MinDist_re	-	2.99	
	MinDist_nr	-	0.93		Slo_mean	-	2.85	
	For_crops	+	0.86		Rural_fam	-	2.55	
	Income	-	0.85		Min_comp	-	2.51	
	MinDist_re	-	0.83		MinDist_ro	-	2.31	
	Prot_areas	-	0.83		Rock	-	2.21	
	Urb_pop	+	0.83		Annu_crops	-	2.11	
ENN[e]	Alt_mean	-	9.72	36.83	Alt_mean	-	4.35	33.1
	Slo_mean	-	3.75		Slo_mean	-	3.91	
	MinDist_nr	+	2.87		Income	-	3.91	
	MinDist_sm	+	2.76		Law_enf	-	3.87	
	Slo_mean_d	-	2.56		Rural_fam	-	3.87	
	Urb_pop	-	2.38		MinDist_re	-	3.87	
	Income	-	2.24		Prot_areas	-	3.68	
	MinDist_re	-	1.97		Min_comp	-	3.56	
	Perm_crops	+	1.76		GPM	-	2.50	
	ICMS_eco	-	1.58					
	Rural_fam	-	1.57					
	Law_enf	-	1.46					
	Roads_den	-	1.21					
	Cov_areas	-	1.00					

"Table 2.3, conclusion."

	RF				STEP			
	Variables	**+/-[a]**	**%imp[b]**	**%var[c]**	**Variables**	**+/-**	**%imp**	**%var**
LSI[f]	Prot_areas	+	1.41	10.57	Rural_fam	-	3.49	36.2
	MinDist_sm	+	1.28		Slo_mean	-	3.49	
	Min_comp	+	1.11		MinDist_sm	-	3.49	
	MinDist_nr	+	0.96		Prot_areas	-	3.49	
	Den_pop	-	0.86		Alt_mean	-	3.49	
	Income	+	0.82		Min_comp	-	3.49	
	Perm_crops	-	0.74		MinDist_nr	-	3.49	
	MinDist_ro	+	0.71		Urb_pop	-	3.35	
	MinDist_re	+	0.65		Cov_areas	-	2.97	
	Soil	+	0.55		Perm_crops	-	2.90	
	Urb_pop	+	0.51		MinDist_re	-	2.55	
	ICMS_eco	+	0.48					
	Rail_den	+	0.48					
PD[g]	Roads_den	+	14.10	59.41	MinDist_ro	-	4.24	39.24
	Slo_mean_d	-	7.02		Slo_mean_d	-	4.24	
	MinDist_ro	+	4.14		Slo_min	-	4.24	
	Alt_mean	+	3.88		MinDist_sm	-	4.11	
	Income	+	3.24		Rural_fam	-	4.11	
	MinDist_sm	+	2.76		Law_enf	-	3.94	
	Den_pop	+	2.73		Rail_den	-	3.73	
	Urb_pop	+	2.72		MinDist_re	-	3.69	
	MinDist_nr	+	2.16		Emp	-	2.84	
	MinDist_ri	-	2.11		Den_pop	+	2.29	
	Rail_den	-	1.82		Urb_pop	-	1.82	
	Annu_crops	+	1.60					
	Perm_crops	-	1.51					
	Min_conc	+	1.38					
	MinDist_re	+	1.38					
	Law_enf	+	1.36					
	Rural_fam	+	1.32					
	ICMS_eco	+	1.26					
	Slo_min	+	1.14					
	Slo_mean	+	0.92					
	Emp	+	0.84					

[a] The symbol "-" indicates the relationship is negatively correlated (low values of independent variable go with high values of dependent variable), while the symbol "+" indicates the relationship is "positively correlated" (high values of independent variable go with high values of dependent variable). [b]Relative importance of each variable measured as a percentage of its contribution to explain the variance (%var).
[c]Percentage of variance explained. For STEP the %var was calculated based on the r^2 value of the final output model, multiplied by 100 to turn it into a percentage value.

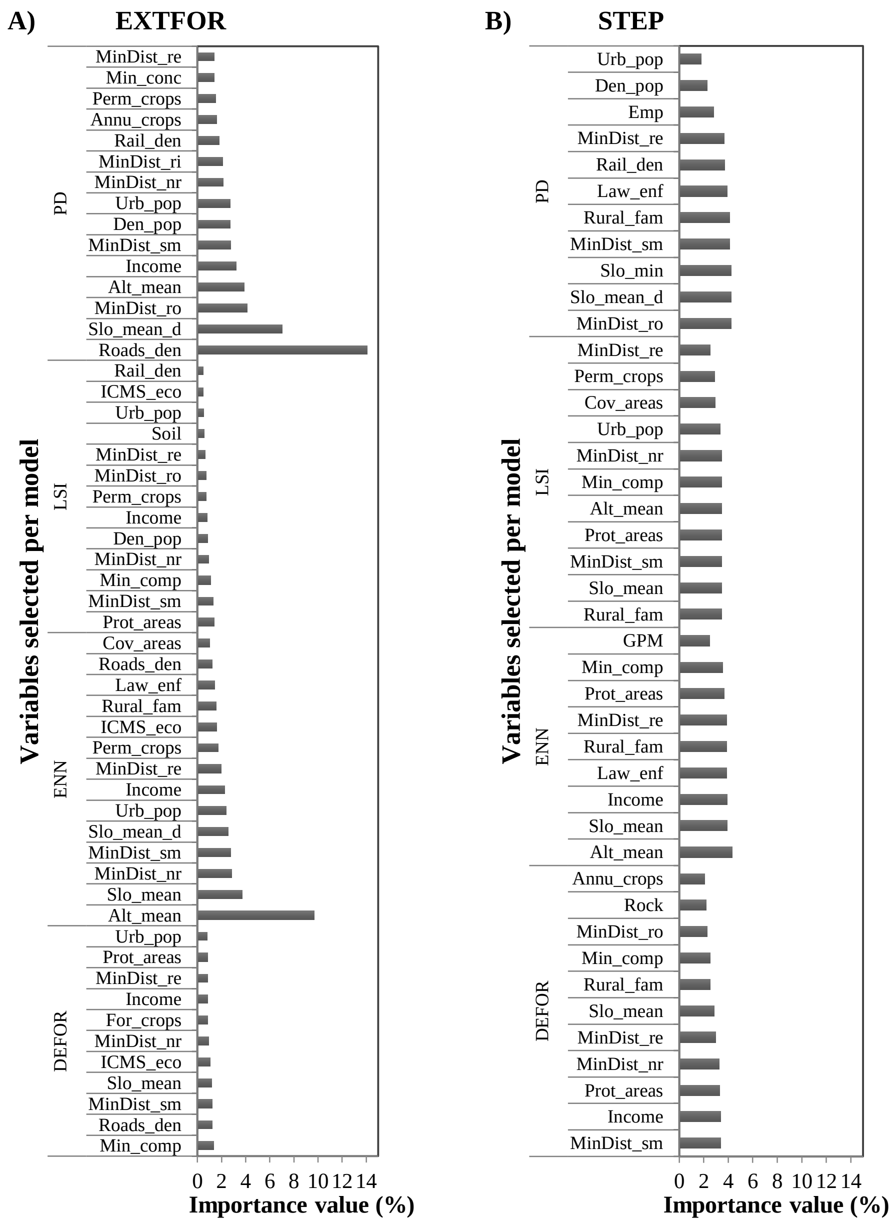


Figure 2.2 Relative importance plots for independent variables from (A) random forest (RF) and (B) stepwise multiple regression (STEP) analyses. See text and Table 2.2 for variable definitions.

Of the four dependent variables, the RF model performed best for patch density (PD), explaining 59.41% of its variance. A large number of independent variables were identified as having some role in explaining PD variations between municipalities; those with the highest importance were associated with the road network (the density of roads, Roads_den, and the minimum distance of forest patches to the nearest road, MinDist_ro) or were topographic (the mean slope in deforested patches, Slo_mean_d, and the mean altitude of each municipality, Alt_mean).

The mean euclidean nearest-neighbour distance between forest patches (ENN) had the second highest amount of its variation explained by RF (36.83%). The mean altitude of each municipality was the most important independent variable here, and other topographic variables (the mean slope across each whole municipality, Slo_mean, and the mean slope within deforested areas, Slo_mean_d) were also relatively important, as were distance factors (the mean distance of forest patches within the municipality to the nearest protected area, MinDist_nr, and the nearest steel mill, MinDist_sm).

In contrast, only 10.57% of variation in the landscape shape index (LSI) was explained by RF. Distance variables were again amongst the most important here, but unlike PD and ENN, natural component independent variables (the amount of protected area in each municipality, Prot_areas, and the index of financial compensation for mineral extraction, Min_comp) are also identified as relatively important for explaining LSI.

RF was also relatively poor at explaining variation in the total amount of deforestation (DEFOR, only 11.23% of variance explained), as it was not linked temporally. No variables have importance measures above 20% here, and the most important ones are a mix of natural components, accessibility, minimum distances and topographic factors.

Overall, five independent variables appeared in the RF models for all four dependent variables: the minimum distance from forest patches to the nearest steel mill (MinDist_sm), the financial compensation for the total amount of protected area (ICMS_eco), the per capita income, the minimum distance to the nearest reservoir (MinDist_re), and the percentage of population categorised as urban (Urb_pop).

2.3.2 Comparisons of RF with STEP

Outcomes from the STEP analysis are shown alongside those for RF, in as comparable a form as possible (Table 2.3, Figure 2.2). Note that, although "percentage importance" values are quoted for both the RF and STEP outputs, these values are not quantitatively comparable between these two methods' outputs. Rather, these values allow us to rank the independent variables in terms of their relative importance for explaining the variability of each dependent variable in each analysis. The percentages of variance explained by the two models are, however, comparable. Both approaches provided evidence of relevant relationships between the dependent variables (deforestation and forest fragmentation metrics) and the independent variables (socio-economic and bio-geophysical factors), but the results are mixed in terms of the independent variables selected as being most important by each analysis.

Like RF, STEP performed best for patch density (PD), but explained less (39.24% c.f. 59.41% for RF) of PD variation between municipalities. There was also a strong similarity between the independent variables selected by the two approaches for the PD models, since all of the independent variables selected in the STEP model were also selected in the RF model. Nine independent variables were found to be of importance by the two approaches; of these, two were topographic (the mean slope of deforestation patches, Slo_mean_d, and the minimum slope within each municipality, Slo_min), three were minimum distances (to the nearest road, MinDist_ro, steel mill, MinDist_sm, and reservoir, MinDist_re), two were human component variables (population density, Den_pop, employment, Emp, and urban population, Urb_pop), and one was a natural component variable (the number of rural family farms, Rural_fam). The density of roads (Roads_den) was the variable identified as being most important by RF, while a similar variable, the minimum distance to the nearest road (MinDist_ro) had the highest importance in the STEP model.

The STEP analysis also followed RF in finding the second highest value of explained variance for ENN (36.83% variance explained by RF, 33.10% by STEP). Six independent variables were found by both approaches to be important for explaining variance in ENN, namely mean altitude (Alt_mean), income (Income), the law enforcement capacity (Law_enf), the minimum distance to the closest reservoir (MinDist_re), the number of rural family farms (Rural_fam), and the mean slope of each municipality (Slo_mean). As for RF, the mean altitude had the highest value of importance in the STEP analysis.

The STEP analysis performed relatively well (36.20% of variance explained) for the landscape shape index (LSI). This is in contrast to RF, which performed worst for this variable (10.57% of variance explained). However, similar independent variables were selected by both statistical approaches, there being seven common to them: three were minimum distances (to the nearest protected area, Min_Dist_nr, reservoir, Min_Dist_re, and steel mill, Min_Dist_sm), three others were natural components (the amount of permanent crops, Perm_crops, the amount of protected area, Prot_areas, and the financial compensation for mineral extraction, Min_comp), and the seventh was a human component variable (urban population, Urb_pop). The amount of protected area was the most important variable for RF, while there were seven equally important variables in the STEP models (Table 2.3).

The STEP analysis also performed relatively well (30.90% of variance explained) in modelling the variation of the total amount of deforestation (DEFOR), compared to RF (11.23% of variance explained). Seven independent variables were found to be of importance in the outputs of both approaches: three minimum distances (the minimum distance to the nearest protected area, MinDist_nr, steel mill, MinDist_sm, and reservoir, Min_Dis_re), two natural components (financial compensation for mineral extraction, Min_comp, and the amount of protected area, Prot_area) one human component (income), and one topographic variable (the mean slope within each municipality, Slo_mean). The most important variables were financial compensation for mineral extraction (Min_comp) for RF, and the minimum distance to the nearest steel mill (MinDist_sm) for STEP.

A single common independent variable, the minimum distance to the nearest reservoir (Min_Dist_re) was selected in the models for all of the dependent variables by both statistical approaches. Whereas five independent variables were common across the four RF models (see above), only one was common across all of the STEP models (the minimum distance to the nearest reservoir (MinDist_re). Notwithstanding these differences, we identified a strong similarity between the independent variables selected in the corresponding RF and STEP models.

2. 4 Discussion

2.4.1 Random forest analysis

In the RF models' outputs, we observed that there are some strong relationships between the the socio-economic and bio-geophysical parameters (our independent variables) and deforestation and forest fragmentation metrics (our dependent variables). RF performed best for patch density (PD), explaining almost 60% of its variance – a relatively high value for ecological studies. It also performed relatively well for patch isolation (mean Euclidean nearest neighbour distance – ENN), explaining 36.83% of its variance, but relatively poorly for the landscape shape index (LSI) and the deforestation (DEFOR), explaining only 10.57% and 11.23% of their variance, respectively. In terms of model performance, this may suggest that the random forest approach is good at identifying parameters that describe the distribution of patches within a landscape (their density and mean separation from each other), but weaker at describing both more macro-scale factors (the overall amount of deforestation) and more micro-scale factors (the shapes of the individual patches). Alternatively, these results could be interpreted as indicating that the patch-distribution scale variables (PD and ENN) are more closely linked to the independent variables we have considered here than are either the bulk amount of deforestation (DEFOR) or the patch shape (LSI). It is important to mention that, even using a very large dataset comprising many independent variables, much of the variance in all of the dependent variables was not accounted by our models. In addition, the question of whether it is primarily the nature of the model or the nature of the independent variables that has led to this finding is not answerable by this first application of RF to this type of data, and remains to be addressed by further investigation.

Turning now to consideration of the independent variables, we found that some of them were particularly strongly related to some of the dependent variables, for example road density and PD, and mean altitude and ENN. However, neither the nature of, nor the reason (i.e. whether they are causatively-linked or simply co-vary) for these links are elucidated by RF. Despite these cases of strong individual-variable links, no single independent variable was found to be related to all of the dependent variables. Geist and Lambin (2002), who investigated the causes of deforestation of tropical forests, also did not find a single important factor. They concluded that forest loss is due to a combination of factors that vary with historical and geographical context. We conclude from the present study that we can expect the same for forest fragmentation.

At the level of independent variable categories, we found that those from the Distance, Accessibility, Natural, and Topography categories contributed most to explaining variance in the RF outputs. Variables from the Human component made less contribution, and variables from the Productivity and Institutional components made hardly any contribution. Additionally, we found that independent variables from Accessibility, Natural, and Topography categories were the most-important independent variable explaining each dependent variable in RF models. This suggests that the physical environment is more important to determine variations in forest fragmentation between municipalities, than social or economic issues. This finding has important implications for management policies aimed at conserving the Atlantic forest, and possibly other biomes that are fragmenting under anthropogenic pressures, although it requires further evidence to be confirmed. Thus, although this ordering of importance of the different types of variables is quite coherent across the RF model outputs, the question remains as to whether it is "true". Claims to this effect are supported by noting that variables that random forest-type methods have identified as most important for classification have been found to coincide with ecological expectations in the literature (Cutler et al. 2007, Wei et al. 2010, Ellis et al. 2012).

2.4.2 Comparisons of RF with STEP

Like RF, the STEP analyses found some strong relationships between the socio-economic and bio-geophysical independent variables, and the deforestation and forest fragmentation dependent variables. STEP followed RF in finding the most explained variance and strongest relationships for PD. STEP also followed RF in finding the second highest explained variance for ENN. Unlike RF, however, there was less difference in the performances of STEP across the four dependent variables: while the explained variances from RF ranged from 10.57% to 59.41%, STEP explained between 30 and 40% of the variance of all four dependent variables.

Overall, there was more agreement than disagreement in terms of the selection and importance of the independent variables between the two approaches. A reasonable number of independent variables was selected as important and shared by them. Considering the categories of independent variables, both approaches found that variables from the Productivity and Institutional components were of little importance, and variables from the Human and Accessibility components were of intermediate importance. Variables from the Distance, Natural and Topography categories were considered of higher importance in models from both methods. In the STEP models, we found that the most-important independent variable explaining each dependent variable model also belonged to those categories.

Only one variable was selected in all deforestation and fragmentation models by both statistical approaches: the minimum distance to the nearest reservoir (MinDist_re). This finding reflects the existence of a large number of reservoirs in the state of Minas Gerais, and the resulting high level of channel fragmentation of important river systems, such as the São Francisco and Paraiba do Sul (Nilsson et al. 2005).

Notwithstanding these similarities between the outcomes of the two modelling approaches, differences between them are evident. However, the reasons for these differences are not clear from our results, and require further investigation. Nonetheless, in theory, one would expect the RF outputs to identify more reliably than STEP the factors that have greatest influence over deforestation and forest fragmentation. This expectation arises from the greater robustness of random-forest type methods compared to traditional regression approaches.

Unlike traditional regression, which has well known weaknesses, despite still being widely used in ecology (Whittingham et al. 2006), random forest methods make no assumptions about the distributions of variables and are robust to outliers in predictor variables. They can also handle situations where the number of predictor variables exceeds the number of observations and have a novel variable importance measure, which does not suffer the shortcomings of traditional variable selection methods, such as selecting only one or two variables among a group of equally good but highly correlated predictors (Cutler et al. 2007). Thus, the greater range of values of explained variance in the RF outputs compared to the STEP outputs, may be indicative of their greater robustness and ability to distinguish meaningfulness relationships. Furthermore, many studies that have applied classical regression approaches to understand the drivers of forest cover changes (e.g. Jaimes et al. 2010,Gao and Li 2011, Freitas et al. 2013, Gong et al. 2013) may have had to use a restricted number of independent variables to be able to satisfy requirements of normality, which could have hindered the analyses, whereas the flexibility and robustness of RF overcomes such limitations.

However, despite its advantages, RF has limitations. The main one is that, unlike traditional regression methods, it does not produce relationships between independent and dependent variables that have simple representations (such as linear equations), and this can make ecological interpretation difficult (Cutler et al. 2007). Therefore, the RF outcomes, and the inferences we make from them, cannot be converted into equations for quantitatively predicting changes in the deforestation and forest fragmentation metrics that might arise from changes in the bio-geophysical and socio-economic variables considered here. That is not the intention here, and we argue that it would be over-simplistic to expect that predictive equations of this kind would be at all useful. Instead, RF has exploited structure in our high-dimensional data set not "visible" to STEP in the PD and ENN models to provide an apparently clearer picture of these metrics' relationships to the independent variables. In the cases of the LSI and DEFOR models, we infer from the poorer performance of RF compared to the STEP method that this exploitation of additional structure can reduce the explanatory power of the models produced in some situations, whilst improving it in others.

2.5 Conclusion

Understanding spatial relationships between patterns of deforestation and forest fragmentation and socio-economic and bio-geophysical factors is important for land use management. The main contribution of this study is the testing of a relatively new method for detecting this kind of relationship (RF), its application to a very large dataset, and its comparison with a traditional multiple linear regression method. We found that RF performs better than multiple regression at explaining metrics describing forest patch patterns (PD and ENN), while it appears to be less capable of capturing the variations of metrics describing a both broader landscape structure (DEFOR) and finer, patch-scale characteristics (LSI). The reasons for these differences in performance remain uncertain. However, given the well-established advantages of decision-tree-based methods over those of classical multiple regression, we suggest that the reasons for these differences are likely to be because the patch-pattern metrics vary in less smooth or monotonic ways – ways that RF is able to capture, but multiple regression is not. In contrast, DEFOR and LSI may vary more smoothly and monotonically in relation to the independent variables, i.e. in ways that traditional regression methods are able to pick up more easily, although why the performance of RF for these variables is relatively weak remains unexplained. Nevertheless, we have shown that RF provides a promising methodology for identifying these relationships, and that it has the potential to be an effective tool for providing essential information for aiding land use management decisions, not only in terms of planning, but also for conservation actions, as proposed by Zanella et al. (2012), in cases of high rates of anthropogenic biodiversity loss.

The initial investigation reported in the present study is, however, only a first step in exploiting this method's potential. One aspect that requires further consideration is the scale of the study area and the very wide variety of socio-economic and bio-geophysical contexts, which it encompasses. Even in relatively small areas, a multitude of diverse factors are at work (Qasim et al. 2013), and variations in contexts may have influenced model performance in the present study. Landscape pattern is scale-sensitive (Gao and Li 2011) and the unusually large degree of heterogeneity in the Atlantic forest biome is likely only to exacerbate this issue. Policies need to be crafted at appropriate spatial scales and with specific contexts in mind. Thus, an important development of this initial study of RF application to cases of deforestation and forest fragmentation would be to repeat it at different spatial scales, to identify more precisely the socio-economic and bio-geophysical factors associated with these processes.

Acknowledgements

The authors would like to thank the Federal University of Lavras (UFLA) for providing the data. L. Zanella would like to acknowledge support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), who provided a PhD scholarship, and Ricardo Solar and Teotônio Soares de Carvalho for assistance in the statistical analysis.

REFERENCES

Barton K. 2014. Package "MuMIn" Available from http://cranr-projectorg/web/packages/MuMIn/MuMInpdf (Accessed August 2014).

Beilin R, et al. 2014. Analysing how drivers of agricultural land abandonment affect biodiversity and cultural landscapes using case studies from Scandinavia, Iberia and Oceania Land use policy 36:60–72.

Bonilla-Moheno M, Mitchell Aide T, Clark ML. 2012 The influence of socioeconomic, environmental and demographic factors on municipality scale land-cover change in Mexico Regional Environmental Change 12:543–557.

Breiman L, Friedman JH, Olshen RA, Stone CI. 1984. Classification and regression trees. Belmont Calif.: Wadsworth.

Breiman L. 2001. Random forests. Mach Learn 45:5–32.

Burnham K, Anderson D (2002) Model selection and multimodel inference: a practical information-theoretic approach Springer-Verlag, New York.

Carvalho LMT, Scolforo JRS. 2008. Inventário Florestal de Minas Gerais: Monitoramento da Flora Nativa 2005-2007. Lavras, Editora: UFLA, 357p.

Carvalho LMT, Scolforo JRS. Unpublished data. Inventário Florestal de Minas Gerais: Monitoramento da Flora Nativa 2007-2009; 2009-2011.

Cutler A, Stevens J. 2006. Random Forests for Microarrays In: Kimmel A, Oliver B (eds). Methods in Enzymology Academic Press, San Diego, pp 422–432.

Cutler DR, Edwards TC, Beard KH, et al. 2007. Random forests for classification in ecology Ecology 88:2783–92.

Efroymson M. 1960. Multiple regression analysis. In: A Ralston and HS Wilf (eds) Mathematical methods for digital computers: New York, John Wiley, pp 191–203.

Ellis EC. 2011. Anthropogenic transformation of the terrestrial biosphere Phil Trans R Soc A 369:1010–1035.

Freitas MWD de, Santos JR dos, Alves DS. 2013. Land-use and land-cover change processes in the Upper Uruguay Basin: linking environmental and socioeconomic variables Landsc Ecol 28:311-327.

Furundzic D. 1998. Application example of neural networks for time series analysis: rainfall-runoff modelling Signal Processing 64:383–396.

Gao J, Li S. 2011. Detecting spatially non-stationary and scale-dependent relationships between urban landscape fragmentation and related factors using Geographically Weighted Regression Appl Geogr 31:292–302.

Geist HJ, Lambin EF. 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. BioScience. 52 (2): 143-150.

Gilbert A, Chakraborty J. 2011. Using geographically weighted regression for environmental justice analysis: Cumulative cancer risks from air toxics in Florida Soc Sci Res 40:273–286.

Gong C, Yu S, Joesting H, Chen J. 2013. Determining socioeconomic drivers of urban forest fragmentation with historical remote sensing images Landsc Urban Plan 117:57–65.

Hocking RR, Mar N. 1976. A Biometrics Invited Paper The Analysis and Selection of Variables in Linear Regression Biometrics 32:1–49.

IBGE – Brazilian Institute of Geography and Statistics. 2015. Estados. Minas Gerais. Available from: http://www.ibge.gov.br/estadosat/perfil.php?sigla=mg. (Accessed May 2015).

IBGE. 2004. Mapa de Biomas e de Vegetação Available from ftp://ftpibgegovbr/Cartas_e_Mapas/Mapas_Murais/ (Accessed August 2014).

Jaimes NBP, Bosque Sendra J, Franco R, et al. 2010. Exploring the driving forces behind deforestation in the state of Mexico (Mexico) using geographically weighted regression Appl Geogr 30:576-591.

James G, Witten D, Hastie T, Tibshirani R. 2013. An introduction to statistical learning Springer, New York.

Kadane JB, Lazar NA. 2004. Methods and criteria for model selection J Am Stat Assoc 99:279–290.

Kleinbaum D, Kupper L, Nizam A, Rosenberg E. 1998. Applied regression analysis and other multivariable methods, 3rd ed Pacific Grove: Duxbury Press.

McGarigal K, Cushman SA, Ene E. 2012. FragStats v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps Computer software program produced by the authors at the University of Massachusetts, Amherst Available from: http://wwwumassedu/landeco/research/fragstats/fragstatshtml (Accessed May 2013).

Nilsson C, Reidy CA, Dynesius M, Revenga C. 2005. Fragmentation and Flow Regulation of the World's Large River Systems. Science 308:405-408.

Parcerisas L, Marull J, Pino J, Tello E, Coll F, Başnou C. 2012. Land use changes, landscape ecology and their socioeconomic driving forces in the Spanish Mediterranean coast (El Maresme County, 1850–2005). Environ Sci Policy 23:120–132.

Prasad AM, Iverson LR, Liaw A. 2006. Newer Tree Classification and Techniques: Forests Random Prediction Bagging for Ecological Regression Ecosystems 9:181–199.

Qasim M, Hubacek K, Termansen M. 2013. Underlying and proximate driving causes of land use change in district Swat, Pakistan Land use policy 34:146–157.

Quezada ML, Arroyo-Rodríguez V, Pérez-Silva E, Aide TM. 2013. Land cover changes in the Lachuá region, Guatemala: patterns, proximate causes, and underlying driving forces over the last 50 years. Reg Environ Chang 14:1139–1149.

Riitters KH. et al. 1995. A factor analysis of landscape pattern and structure metrics. Landsc Ecol 10:23–39.

Scolforo JR, Oliveira AD de, Carvalho LMT. 2008. Zoneamento ecológico-econômico do estado de minas gerais: Componente socioeconômico. Lavras: Editora da UFLA.

Scolforo JRS, Carvalho LMT. 2006. Mapeamento e inventário da flora nativa e dos reflorestamentos em Minas Gerais Lavras: Editora da UFLA 288p.

Smith SJ, Ellis N, Pitcher CR. 2011. Conditional variable importance in R package extendedForest Available from http://gradientforestr-forger-projectorg/Conditional-importancepdf (Accessed January 2014).

Venables WN, Ripley BD. 2002. Modern Applied Statistics with S, 4thed. New York: Springer.

Wei C-L, Rowe GT, Escobar-Briones E, et al. 2010. Global patterns and predictions of seafloor biomass using random forests. PLoS One 5:e15323.

Whittingham MJ, Stephens P a, Bradbury RB, Freckleton RP (2006) Why do we still use stepwise modelling in ecology and behaviour? J Anim Ecol 75:1182–1189.

Zanella L, Borém R, Souza C, Alves HM, Borém FM. 2012. Atlantic Forest Fragmentation Analysis and Landscape Restoration Management Scenarios. Nat Conservação 10:57–63.

APPENDIX

Supplementary material

List S1 Literature consulted to determine commonly used landscape metrics

Arroyo-Mora, J. P., Sánchez-Azofeifa, G. A., Rivard, B., Calvo, J. C., & Janzen, D. H. (2005). Dynamics in landscape structure and composition for the Chorotega region, Costa Rica from 1960 to 2000. *Agriculture, Ecosystems & Environment*, *106*(1), 27–39. doi:10.1016/j.agee.2004.07.002

Bailey, D., Billeter, R., Aviron, S., Schweiger, O., & Herzog, F. (2006). The influence of thematic resolution on metric selection for biodiversity monitoring in agricultural landscapes. *Landscape Ecology*, *22*(3), 461–473. doi:10.1007/s10980-006-9035-9

Banks-Leite, C., Ewers, R. M., & Metzger, J. P. (2013). The confounded effects of habitat disturbance at the local, patch and landscape scale on understorey birds of the Atlantic Forest: Implications for the development of landscape-based indicators. *Ecological Indicators*, *31*, 82–88. doi:10.1016/j.ecolind.2012.04.015

Batista, T., & Mendes, P. (2012). Suitable methods for landscape evaluation and valorization: the third dimension in landscape metrics. *Acta Botanica …*, (August), 37–41. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/12538078.2012.696930

Brown, G. G., & Reed, P. (2013). Social Landscape Metrics : Measures for Understanding Place Values from Public Participation Geographic Information Social Landscape Metrics : Measures for Understanding Place Values from Public Participation Geographic Information Systems (PPGIS). *Landscape Research*, *37:1*(June 2013), 73–90.

Cushman, S. a., McGarigal, K., & Neel, M. C. (2008). Parsimony in landscape metrics: Strength, universality, and consistency. *Ecological Indicators*, *8*(5), 691–703. doi:10.1016/j.ecolind.2007.12.002

Fichera, C. R. (2012). Land Cover classification and change-detection analysis using multi-temporal remote sensed imagery and landscape metrics. *European Journal of Remote Sensing*, 1–18. doi:10.5721/EuJRS20124501

Frank, S., Fürst, C., Koschke, L., & Makeschin, F. (2012). A contribution towards a transfer of the ecosystem service concept to landscape planning using landscape metrics. *Ecological Indicators*, *21*, 30–38. doi:10.1016/j.ecolind.2011.04.027

Hassett, E. M., Stehman, S. V., & Wickham, J. D. (2011). Estimating landscape pattern metrics from a sample of land cover. *Landscape Ecology*, *27*(1), 133–149. doi:10.1007/s10980-011-9657-4

Hickey, J. R., Carroll, J. P., & Nibbelink, N. P. (2012). Applying Landscape Metrics to Characterize Potential Habitat of Bonobos (Pan paniscus) in the Maringa-Lopori-Wamba Landscape, Democratic Republic of Congo. *International Journal of Primatology*, *33*(2), 381–400. doi:10.1007/s10764-012-9581-8

Hill, M. O., Roy, D. B., & Thompson, K. (2002). Hemeroby, urbanity and ruderality: bioindicators of disturbance and human impact. *Journal of Applied Ecology*, *39*(5), 708–720. doi:10.1046/j.1365-2664.2002.00746.x

Hou, W., & Walz, U. (2013). Enhanced analysis of landscape structure : Inclusion of transition zones and small-scale landscape elements. *Ecological Indicators*, *31*, 15–24. doi:10.1016/j.ecolind.2012.11.014

Kilgore, M. a., Snyder, S. a., Block-Torgerson, K., & Taff, S. J. (2013). Challenges in characterizing a parcelized forest landscape: Why metric, scale, threshold, and definitions matter. *Landscape and Urban Planning*, *110*, 36–47. doi:10.1016/j.landurbplan.2012.09.009

Kupfer, J. a. (2012). Landscape ecology and biogeography: Rethinking landscape metrics in a post-FragsStats landscape. *Progress in Physical Geography*, *36*(3), 400–420. doi:10.1177/0309133312439594

Li, X., Lin, J., Chen, Y., Liu, X., & Ai, B. (2013). Calibrating cellular automata based on landscape metrics by using genetic algorithms. *International Journal of Geographical Information Science*, *27:3*(June 2013), 594–613.

Liu, D., Hao, S., Liu, X., Li, B., He, S., & Warrington, D. N. (2012). Effects of land use classification on landscape metrics based on remote sensing and GIS. *Environmental Earth Sciences*, *68*(8), 2229–2237. doi:10.1007/s12665-012-1905-7

McGarigal, K., Tagil, S., & Cushman, S. a. (2009). Surface metrics: an alternative to patch metrics for the quantification of landscape structure. *Landscape Ecology*, *24*(3), 433–450. doi:10.1007/s10980-009-9327-y

Neel, M. C., McGarigal, K., & Cushman, S. a. (2004). Behavior of class-level landscape metrics across gradients of class aggregation and area. *Landscape Ecology*, *19*(4), 435–455. doi:10.1023/B:LAND.0000030521.19856.cb

Ozdemir, I., Mert, A., & Senturk, O. (2012). Predicting Landscape Structural Metrics Using Aster Satellite Data. *Journal of Environmental Engineering and Landscape Management*, *20*(2), 168–176. doi:10.3846/16486897.2012.688371

Parcerisas, L., Marull, J., Pino, J., Tello, E., Coll, F., & Basnou, C. (2012). Land use changes, landscape ecology and their socioeconomic driving forces in the Spanish Mediterranean coast (El Maresme County, 1850–2005). *Environmental Science & Policy*, *23*, 120–132. doi:10.1016/j.envsci.2012.08.002

Pardini, R., Bueno, A. D. A., & Gardner, T. (2010). Beyond the fragmentation threshold hypothesis: regime shifts in biodiversity across fragmented landscapes. *Plos One*, *5*(10). doi:10.1371/journal.pone.0013666

Ramezani, H., Holm, S., Allard, A., & Ståhl, G. (2013). Norsk Geografisk Tidsskrift - Norwegian Journal of Geography A review of sampling-based approaches for estimating landscape metrics A review of sampling-based approaches for estimating landscape metrics. *Norsk Geografisk Tidsskrift - Norwegian Journal of Geography*, *67:2*(June), 61–71.

Riitters, K. H., Neil, R. V. O., Hunsaker, C. T., Wickham, J. D., Yankee, D. H., & Timmins, S. P. (1995). A factor analysis of landscape pattern and structure metrics. *Landscape Ecology*, *10*(1), 23–39.

Schindler, S., von Wehrden, H., Poirazidis, K., Wrbka, T., & Kati, V. (2013). Multiscale performance of landscape metrics as indicators of species richness of plants, insects and vertebrates. *Ecological Indicators*, *31*, 41–48. doi:10.1016/j.ecolind.2012.04.012

Sundell-Turner, N. M., & Rodewald, A. D. (2008). A comparison of landscape metrics for conservation planning. *Landscape and Urban Planning*, *86*(3-4), 219–225. doi:10.1016/j.landurbplan.2008.03.001

Syrbe, R.-U., & Walz, U. (2012). Spatial indicators for the assessment of ecosystem services: Providing, benefiting and connecting areas and landscape metrics. *Ecological Indicators*, *21*, 80–88. doi:10.1016/j.ecolind.2012.02.013

Szabó, S., Csorba, P., & Szilassi, P. (2012). Tools for landscape ecological planning—Scale, and aggregation sensitivity of the contagion type landscape metric indices. *Carpathian Journal of Earth and …*, *7*(3), 127–136. Retrieved from http://www.researchgate.net/publication/230703111_TOOLS_FOR_LANDSCAPE_ECOLOGICAL_PLANNING__SCALE_AND_AGGREGATION_SENSITIVITY_OF_THE_CONTAGION_TYPE_LANDSCAPE_METRIC_INDICES/file/9fcfd5033762eb1143.pdf

Uuemaa, E., Antrop, M., & Marja, R. (2009). Landscape Metrics and Indices: An Overview of Their Use in Landscape Research Imprint/Terms of Use. *Landscape*, 1–28. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Landscape+Metrics+and+Indices+:+An+Overview+of+Their+Use+in+Landscape+Research+Imprint+/+Terms+of+Use#7

Uuemaa, E., Mander, Ü., & Marja, R. (2013). Trends in the use of landscape spatial metrics as landscape indicators: A review. *Ecological Indicators*, *28*, 100–106. doi:10.1016/j.ecolind.2012.07.018

Wu, M. Y., Xue, L., Jin, W. B., Xiong, Q. X., Ai, T. C., & Li, B. L. (2012). Modelling the Linkage Between Landscape Metrics and Water Quality Indices of Hydrological Units in Sihu Basin, Hubei Province, China: An Allometric Model. *Procedia Environmental Sciences*, *13*(2011), 2131–2145. doi:10.1016/j.proenv.2012.01.202

Zaragozí, B., Belda, a., Linares, J., Martínez-Pérez, J. E., Navarro, J. T., & Esparza, J. (2012). A free and open source programming library for landscape metrics calculations. *Environmental Modelling & Software*, *31*, 131–140. doi:10.1016/j.envsoft.2011.10.009

Zhang, N., & Li, H. (2013). Sensitivity and effectiveness and of landscape metric scalograms in determining the characteristic scale of a hierarchically structured landscape. *Landscape Ecology*, *28*(2), 343–363. doi:10.1007/s10980-012-9837-x

Zhang, Y., Odeh, I. O. a., & Ramadan, E. (2013). Assessment of land surface temperature in relation to landscape metrics and fractional vegetation cover in an urban/peri-urban region using Landsat data. *International Journal of Remote Sensing*, *34*(1), 168–189. doi:10.1080/01431161.2012.712227

Zhang, Z., Van Coillie, F., De Clercq, E. M., Ou, X., & De Wulf, R. (2013). Mountain vegetation change quantification using surface landscape metrics in Lancang watershed, China. *Ecological Indicators*, *31*, 49–58. doi:10.1016/j.ecolind.2012.11.013

Zhiming, Z., Van Coillie, F., De Wulf, R., De Clercq, E. M., & Xiaokun, O. (2012). Comparison of Surface and Planimetric Landscape Metrics for Mountainous Land Cover Pattern Quantification in Lancang Watershed, China. *Mountain Research and Development*, *32*(2), 213–225. doi:10.1659/MRD-JOURNAL-D-10-00119.1

CHAPTER 3

MULTI-SCALE RANDOM FOREST ANALYSIS FOR MODELLING RELATIONSHIPS BETWEEN LANDSCAPE PATTERN AND ASSOCIATED FACTORS

Publication status: *In prep.* for submission to *Landscape Ecology*

ABSTRACT

Deforestation and forest fragmentation are considered to be major threats to biodiversity in the tropics. The Brazilian Atlantic Forest (Minas Gerais state, Brazil), which is located within the tropics at 14-23° S, is a hotspot of biodiversity and is severely threatened by anthropogenic deforestation. This study uses a multi-scale approach to investigate deforestation and forest fragmentation in the Brazilian Atlantic Forest and to elucidate the relationships between these processes and a wide variety of socio-economic and bio-geophysical factors. We considered these relationships at the whole biome scale, and at regional and sub-regional scales within the biome. We used a recently developed, machine-learning technique, Random Forests (RF) analysis. We found that 1.69% of the current remaining area of Atlantic Forest in Minas Gerais was lost between 2003 and 2011. We also verified that RF analysis can be used effectively in this context, providing explanations of 67% of variance in deforestation and forest fragmentation metrics. We found that the relationships between these metrics and the socio-economic and bio-geophysical factors used in the models varied from place to place and across spatial scales, and that some metrics were better explained by the RF models at the largest (biome) scale and others at the smallest (sub-regional) scale. Road density emerged as the factor that appeared most commonly in the models explaining variance in the metrics and, in general, factors describing the spatial distribution of the natural, agricultural and infrastructural elements of the landscape occurred more commonly in the models that those describing patterns of population, employment and legal institutions. Given that previous studies identify social and economic factors as being important determinants of landscape structure, we infer that the relative lack of importance of variables of this sort we identify may be due to asynchronous relationships between socio-economic drivers and patterns of deforestation and forest fragmentation.

Keywords: Socio-economic and bio-geophysical factors. Land use and land cover change. Random Forest Regression. Machine-learning technique. Landscape metrics. Deforestation. Forest fragmentation. Brazilian Atlantic Forest. Tropical forests. Minas Gerais State.

3.1 Introduction

Anthropogenic activities have been triggering unprecedented land use and cover changes (LUCC) throughout the world in recent decades (Lambin and Geist 2006). It is estimated that 13 million ha/year of forest were altered worldwide between 1990 and 2005 (Food and Agriculture Organization of The United Nations - FAO 2008). Deforestation and forest fragmentation in the tropics have been reported to be major threats to biodiversity conservation (Brooks et al. 2006), affecting irreversibly a large number of key ecosystem services (Musaoglu et al. 2005, Heistermann et al. 2006). The Brazilian Atlantic Forest, a highly diverse tropical forest (Mittermeier et al. 2005), is a primary example of a biome that has suffered the consequences of anthropogenic activities. This biome formerly covered 1.5 million km^2 along the Brazilian coast (Galindo-Leal and Câmara 2003, Ribeiro et al. 2009), but the current amount of its remaining area is critically small; estimates vary between 7-8% (*SOS Mata Atlântica* and INPE 2008) and 12% (Ribeiro et al. 2009) of its extent at the time of the arrival of the first Europeans in South America in the 16th century. The Atlantic Forest is also considered to be a biodiversity hotspot (Mittermeier et al. 2011, Zachos and Habel, 2011), and one of the most threatened tropical forests in the world.

LUCC is considered to be the result of complex interactions among social, economic, and environmental factors (Lambin et al. 2001, Geist and Lambin 2002). Socio-economic processes are thought to be the primary factors associated with LUCC, which in turn determine the structure, function, and dynamics of most landscapes, so that changes in social structures and processes lead to alterations in the environment (Lorenzoni et al. 2000, Wang and Zhang 2001, Wu and Hobbs 2002). It is also recognised that underlying bio-geophysical factors have also to be taken into account when attempting to understanding the nature of LUCC (Turner II et al. 1995). Hence, conservation actions in the Atlantic Forest need to be formulated on the basis of a clear understanding of the ways in which socio-economic and bio-geophysical factors determine the spatial structure of LUCC. Understanding these relationships is difficult, because of the multiplicity of potentially important factors, and the ways in which they act differently at different scales of space and time (Lambin et al. 2001, Geist and Lambin 2002). Thus, in attempting to make reliable inferences about relationships between these factors and LUCC, it is necessary to account carefully for scale variations, as well as selecting appropriate statistical approaches.

Several previous studies have attempted to explore LUCC at different scales. While large scale studies have tended to report negative LUCC patterns (e.g., loss of forest cover in parts of Mexico (García-Barrios et al. 2009), and in the Brazilian Amazon (Deng et al. 2014)), some local studies within regions have reported positive trends (e.g., increase in forest cover in specific Mexican municipalities (Bonilla-Moheno et al. 2012), and forest transitions in three 10,000ha plots in the Brazilian Atlantic Forest of São Paulo (Lira et al. 2012). Fully understanding and replicating such local studies can help place local trends into a regional context. Local-scale analyses are needed to identify the proximate (e.g., agriculture or pastureland expansion) and underlying (e.g., socioeconomic or demographic) factors driving deforestation and fragmentation (Geist and Lambin 2002, Lambin et al. 2003, Scrieciu 2007). However, local-scale analyses are limited because they are usually geographically constrained and may include only a few case studies, so their results cannot easily be generalized to larger spatial scales. Thus, studies that integrate analyses at different spatial scales are needed. Multi-scale approaches have not yet been implemented in studies of the effects of socio-economic and bio-geophysical factors on deforestation and fragmentation in the Brazilian Atlantic Forest.

This paper attempts to fill this gap in knowledge and to elucidate the relationships between deforestation and forest fragmentation and a very wide variety of socio-economic and bio-geophysical factors at multiple spatial scales in the Brazilian Atlantic Forest of Minas Gerais. Our main objectives are to answer the following questions: (1) How well can deforestation and forest fragmentation patterns be explained by models based on socio-economic and bio-geophysical factors?; (2) To what extent do these models vary across the study area and at different spatial scales?; (3) Which socio-economic and/or bio-geophysical factors are most important across the study area and at different spatial scales for explaining variations in deforestation/fragmentation?

3.2 Methods

3.2.1 Study area

The state of Minas Gerais is located in South-eastern Brazil, between latitudes 14º 03' 28" S and 23º 07' 02" S and longitudes 51º 07' 02" W and 39º 49' 58" W. It covers an area of 58,652,212 ha and is split into 853 municipalities, ranging in area from 285 ha to 1,071,696 ha. It has three biomes within its limits: Cerrado, Caatinga and Atlantic Forest (IBGE 2004). The area with which the present study is concerned comprises the 518 municipalities that fall entirely within the largest contiguous area of the Atlantic Forest biome, and encompasses 34% (19,904,146 ha – total area of the Atlantic Forest in the state covers 41% (24,047,660 ha)) of Minas Gerais (Figure 3.1) (IBGE 2015). This study site was chosen for the purposes of this project, as there is a wide variability across the municipalities in terms of deforestation/fragmentation and in the socio-economic/bio-geophysical variables. In terms of data availability, Minas Gerais is one of the few Brazilian states that has estimates of both forest cover change, and most of the socio-economic/bio-geophysical variables available at the municipality scale.

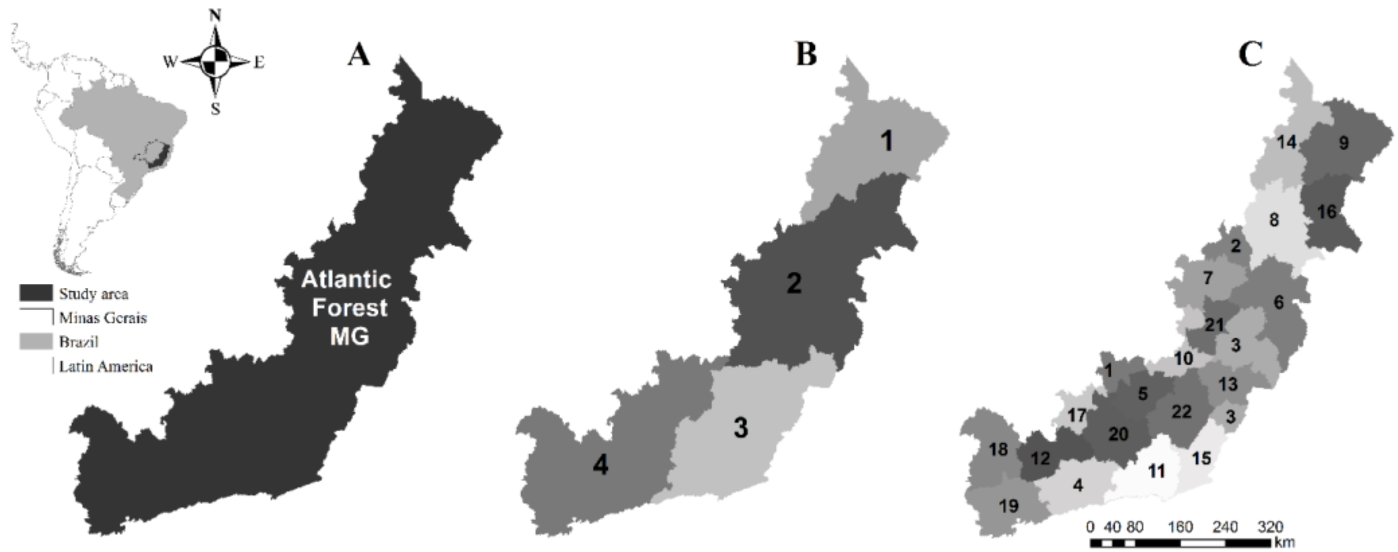


Figure 3.1 Multiple spatial scales considered in this study: A) the Atlantic Forest Biome (Minas Gerais, Bra); B) Regions; C) Sub-regions.

3.2.2 Multiple spatial scales: grouping units

The available demographic and socio-economic data were derived within political administrative units at the scale of municipalities, the smallest administrative units in Brazil. Neighbouring municipalities share a number of potential drivers of land-cover change (e.g., political and economic programs, environmental conditions, social and demographic contexts), which make them relatively easy to group and thus a useful unit for multi-scale land change analyses (Bonilla-Moheno et al. 2012). In our multi-scale approach, the municipalities were grouped into spatial units at different spatial scales as follows. Three scales of municipalities groupings were used: i) sub-regions, our finest scale; ii) regions, the intermediate scale; and iii) biome, the largest scale. Sub-region delimitation was based on administrative groupings of municipalities established by the Forest State Institute of Minas Gerais (IEF), which identifies 22 sub-regions. These vary in area from 353,523 ha to 1,543,000 ha and contain between 9 and 44 municipalities. Region delimitation used existing political regional boundaries, since political decisions are taken within these limits, and also because there is a certain coincidence between them and the distribution of bio-geophysical factors (e.g. homogeneity of relief, soil types, geology, hydrological basins, the amount of remaining forest area, as well as historical occupation). We grouped municipalities within four regions. These vary in area from 3,258,197 ha to 6,475,025 ha and contain between 32 and 189 municipalities. At the biome scale level, we considered the entire study area as a single spatial unit, which groups all 518 municipalities within the Atlantic Forest Biome in Minas Gerais. The historical context of the Minas Gerais colonization provides some evidence of the current differences among the regions and sub-regions we studied, starting with the arrival of Portuguese explorers in the 16th century (Carrato 1968) who first reached the southern region of the area known currently as Minas Gerais State (literal translation: *General Mines*), looking for gold and gemstones (Carrato 1968). This can be related to the fact that this southern region (Region 4 in our study, containing Sub-regions 1, 4, 5, 12, 17, 18, 19 and, partially, 20) is currently the most populated in the state (population density 0.75 people per km^2 while in the northeast region it is 0.15 people per km^2; IBGE 2015). In the 19^{th} century, once mining started to decline and was no longer economically sustainable, miners were forced to move to other economic activities or to other areas, expanding the state boundaries (Carrato 1968). Some ex-miners remained in the south and most of them became coffee producers, increasing the regional economy greatly (Carrato 1968). Other ex-miners moved towards the northeast of the state where they found vast uninhabited areas and became mainly cattle ranchers (Carrato 1968). Nowadays, this region is known for its low socio-economic indicators (IBGE 2015), and in our study is divided into Regions 1 (made up of Sub-regions 9 and 14) and 2 (made up of Sub-regions 2, 3, 6, 7, 8, 10, 16, and 21). Towards the east, settlement was driven strongly throughout the 19^{th} century by the expansion of coffee plantations (Carrato 1968) and followed the southeast in developing a strong economy (IBGE 2015). This region is defined as Region 3 in our study, and contains Sub-regions 3, 11, 13, 15, 22 and, partially, 20.

3.2.3 Variable selection

A large dataset provided by two broader-scale projects developed in the state of Minas Gerais, Brazil, was used in this work: the vegetation monitoring system (Scolforo and Carvalho 2006), and the ecologic-economical zoning of Minas Gerais – ZEE-MG (Scolforo et al. 2008). The dataset comprises land-cover maps from 2003 to 2011, deforestation rate estimates, and measurements of more than 300 socio-economic and bio-geophysical variables.

We calculated deforestation rates from forest cover data recorded every two years between 2003 and 2011 (Scolforo and Carvalho 2006, Carvalho and Scolforo 2008, Carvalho and Scolforo – unpublished data), at the whole-biome and regional scales. Deforestation rates at the municipality scale were calculated using radiometric change detection applied to Landsat images. All deforestation metrics were normalized to the remaining forest area within each municipality/region.Table 3.1 shows the deforestation variables used in this study.

Table 3.1 Deforestation metrics including deforestation rates between 2003 and 2011 for every two years (DEFOR0305; DEFOR0705; DEFOR0709; and DEFOR0911); and, the total area deforested between 2003 and 2011 (DEFOR), as a measurement of deforestation.

Abbreviation	**Description (unit)**
DEFOR0305	Area deforested between 2003 and 2005. Hectares
DEFOR0507	Area deforested between 2005 and 2007. Hectares
DEFOR0709	Area deforested between 2007 and 2009. Hectares
DEFOR0911	Area deforested between 2009 and 2011. Hectares
DEFOR	Area deforested between 2003 and 2011. Hectares

We quantified forest fragmentation in the study area by using a land use map from 2011, also provided by the vegetation monitoring system (Scolforo and Carvalho 2006, Carvalho and Scolforo 2008, Carvalho and Scolforo – unpublished data). FragStats 4.0 (McGarigal et al. 2012) was used to calculate a set of 225 landscape metrics for each of the 518 municipalities. These were then passed through a two-stage filtering process to provide a tractable set of dependent variables for use in our analysis. Firstly, we addressed multicollinearity, which can often be found in landscape metrics (Riitters et al. 1995), by identifying metrics that correlated with each other at a significance level of $p < 0.01$. In each case, we discarded one of the variables and retained the other, taking in account their ecological meaning. The selection criteria used to determine which variable to retain were mathematical simplicity and an intuitive judgment of their explanatory power in terms of ecological meaning. Secondly, we chose metrics from the remaining subset that were commonly used in literature found via a search on the Web of Knowledge website (http://wok.mimas.ac.uk/). The search was carried out for papers published from January 2011 to June 2013, using the key-words "landscape metrics" and/or "landscape indices". This search yielded 48 papers, of which four were found, on inspection, to be out of scope, and five were unavailable to us. The papers consulted in the review can be seen in Zanella (*in prep.* – Chapter 2). The result of this filtering process provided four meaningful metrics representing forest fragmentation: the aggregation index (AI), a measure of the degree to which forested areas are clumped together; the connectance index (CONNECT), a measure of the degree to which forested areas are within a user-specified threshold distance of each other; the mean Euclidean nearest-neighbour distance (ENN), a measure of forest patches' separation from each other; and the patch density (PD), a measure of the degree to which the forest is broken up into individual patches (Table 3.2).Figure 3.2 shows the variability of dependent variables across the municipalities.

Socio-economic and bio-geophysical variables were obtained from the ZEE-MG database. The years for which these variables were collected were limited by the availability of information from national agencies, and ranged from 2003 to 2006. Based on data availability, and following Scolforo et al. (2008), socio-economic variables from four categories – production, natural, human and institutional – were used. Variables from a further three categories of bio-geophysical factors – topographic, distance and accessibility – were also selected. This gave an initial list of more than 300 candidate independent variables. Descriptions of how these variables were calculated can be found in Scolforo et al. (2008). From this list, a tractable sub-set of variables was derived using the process for addressing multicollinearity described above. As a result, a total of 34 socio-economic and bio-geophysical variables were selected as independent variables for use in our multi-scale analysis of deforestation and forest fragmentation drivers (Table 3.3). Some of these variables (highlighted with '*' in Table 3.3) were normalised by the municipality area prior to modelling to allow comparison among municipalities.

Table 3.2 Descriptions of deforestation and forest fragmentation metrics used as dependent variables.

Metric	Category	Formulae	Description (unit)[a]
Aggregation index (AI)	Forest aggregation	$AI = \left[\frac{g_{ii}}{max \rightarrow g_{ii}}\right](100$	AI equals the number of like adjacencies involving the corresponding class, divided by the maximum possible number of like adjacencies involving the corresponding class, which is achieved when the class is maximally clumped into a single, compact patch; multiplied by 100 (to convert to a percentage). If A*i* is the area of class $_i$ (in terms of number of cells) and n is the side of a largest integer square smaller than Ai, and m = A*i* - n_2, then the largest number of shared edges for class $_i$, max-g_{ii} will take one of the three forms: max-g_{ii} = 2n(n-1) , when m = 0, or max-g_{ii}= 2n(n-1) + 2m -1, when m = n, or max-g_{ii} = 2n(n-1) + 2m -2, when m > n. Note, because of the design of the metric, like adjacencies are tallied using the single-count method and all landscape boundary edge segments are ignored, even if a border is provided. $0 \leqq AI \leqq 100$ Given any Pi, AI equals 0 when the focal patch type is maximally disaggregated (i.e., when there are no like adjacencies); AI increases as the focal patch type is increasingly aggregated and equals 100 when the patch type is maximally aggregated into a single, compact patch. g_{ii} = number of like adjacencies (joins) between pixels of patch type (class) $_i$ based on the single-count method. max-g_{ii}= maximum number of like adjacencies (joins) between pixels of patch type (class) $_i$ (see below) based on the single-count method. Percent.

"Table 3.2, conclusion."

Metric	Category	Formulae	Description (unit)[a]
Connectance index (CONNECT)	Forest patch connectance	$CONNECT = \frac{\sum_{j=k}^{n} C_{ijk}}{\frac{n_i(n_i-1)}{2}}(100)$	CONNECT equals the number of functional joinings between all patches of the corresponding patch type (sum of c_{ijk} where $c_{ijk} = 0$ if patch j and k are not within the specified distance of each other and $c_{ijk} = 1$ if patch j and k are within the specified distance), divided by the total number of possible joinings between all patches of the corresponding patch type, multiplied by 100 to convert to a percentage. $0 \leqq$ CONNECT $\leqq 100$. CONNECT = 0 when either the focal class consists of a single patch or none of the patches of the focal class are "connected" (i.e., within the user-specified threshold distance of another patch of the same type). CONNECT = 100 when every patch of the focal class is "connected". Connectance is defined on the number of functional joinings between patches of the corresponding patch type, where each pair of patches is either connected or not based on a user-specified distance criterion. Connectance is reported as a percentage of the maximum possible connectance given the number of patches. Connectance can be based on either Euclidean distance or functional distance. c_{ijk} = joining between patch j and k (0 = unjoined, 1 = joined) of the corresponding patch type (i), based on a user specified threshold distance. n_i = number of patches in the landscape of the corresponding patch type (class).Percent.
Mean Euclidean Nearest-Neighbour (ENN)	Forest patch isolation	$\text{ENN} = \frac{\sum_{j=1}^{n} h_{ij}}{n_i}$	ENN equals the mean distance to the nearest neighbouring patch of forest, based on shortest edge-to-edge distance. h_{ij} = distance (m) from patch j to nearest neighbouring patch of the same type (i, in this case forest). n_i = number of patches of cover type i (forest). Meters.
Patch density (PD)	Forest spatial structure	$\text{PD} = \frac{n_i}{A}(10{,}00000)$	Patch density increases with a greater number of patches within a reference area and therefore reflects landscape fragmentation.

a Details can be found in McGarigal et al. (2012).

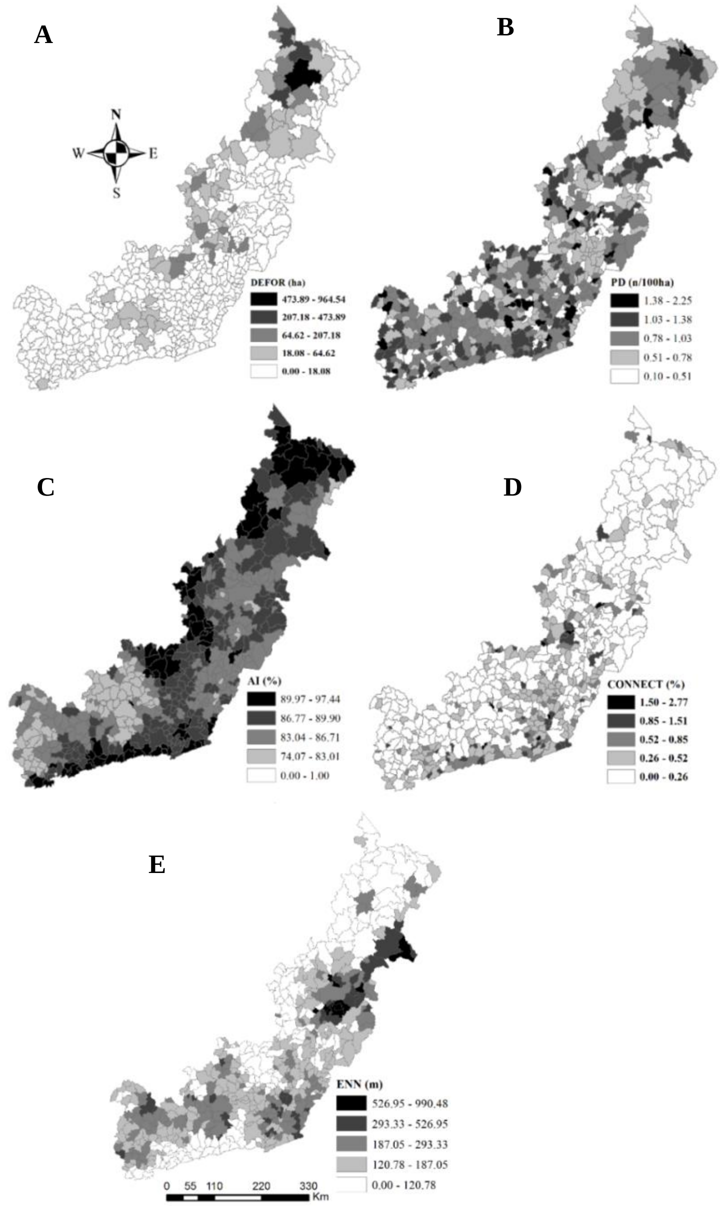


Figure 3.2 Spatial distribution of dependent variable values per municipalities along the Atlantic Forest Biome in Minas Gerais. A) Deforestation (DEFOR); B) Patch density (PD); C) Forest aggregation (AI); D) Patch connectance (CONNECT); E) Forest patch isolation (ENN).

Table 3.3 Socio-economic and bio-geophysical factors used as independent variables.

Category	**Acronym**	**Description[a]**
Productive component	VA_agri	Value added by agriculture sector adjusted by inflation in 2004 (R$[b])
	GGR_GPM	Geometric growth rate of total gross product per municipality (1999-2003)
	GPM	Total gross product per municipality in 2004 (R$)
Natural component	For_crops*	Amount of forest crops per municipality in hectares
	Perm_crops*	Amount of permanent crops per municipality in hectares
	Annu_crops*	Amount of annual crops per municipality in hectares
	Cov_areas*	Amount of area covered by reservoirs per municipality in hectares
	Prot_areas*	Amount of protected areas per municipality in hectares
	Rural_fam*	Number of rural family farms per municipality. Rural family farms are those in which labour is performed by family members only.
	Min_comp	Index of financial compensation for mineral extraction in 2005 (R$)
	ICMS_eco	Financial compensation given to municipalities that have land use restrictions due to protected areas (Ecological ICMS) in 2005 (R$)
	Min_conc	Number of mines in operation by municipality
Human component	Occu_rate	Intensity of usage of available land for economic use, obtained from the total area of the municipality available for economic activities minus protected and covered areas.
	Unemp	Unemployment rate per municipality in 2005
	Emp	Geometric growth rate of formal employment per municipality between 2000 and 2005
	Den_pop	Population density per municipality in 2004
	Urb_pop	Percentage of urban population per municipality in 2004
	Income	Per capita income per municipality (R$ / per capita)
Institutional component	Law_enf	Law Enforcement Capacity measures the ratio between the number of citizens and the number of Military and Civil Policemen, Judges, District Attorneys and Public Defenders in the municipality. It ranges from 0 to 1 (0 - no capacity; 1 - best capacity)

"Table 3.3, conclusion."

Category	Acronym	Description[a]
Topography	Alt_mean	Mean altitude per municipality in meters
	Slo_min	Minimum slope per municipality in degrees
	Slo_mean	Mean slope per municipality in degrees
	Slo_mean_d	Mean slope of areas deforested between 2003 and 2011, per municipality in degrees
	Rock	Predominant rock type covered by forests in the municipality
	Soil	Predominant soil type covered by forests in the municipality
Accessibility	Roads_den	Road density per municipality (km/1.000 km^2)
	Rail_den	Railway density per municipality (km/1.000 km^2)
Distance factors	MinDist_sm	Mean distance of forest patches within the municipality to the closest steel mill in kilometres
	MinDist_ri	Mean distance of forest patches within the municipality to the closest river in kilometres
	MinDist_nr	Mean distance of forest patches within the municipality to the closest protected area in kilometres
	MinDist_ro	Mean distance of forest patches within the municipality to the closest road (highway) in kilometres
	MinDist_re	Mean distance of forest patches within the municipality to the closest reservoir in kilometres
Geographical location	Point_X	Longitude of the centroid of the municipality
	Point_Y	Latitude of the centroid of the municipality

[a] The full description on how the variables were calculated and their description are found in Scolforo et al. (2008).
[b] Brazilian currency
*Variables scaled to the municipality area prior to modelling to allow comparison.

3.2.4 Random Forests analysis

To investigate in detail the relationships between deforestation/forest fragmentation metrics and socio-economic/bio-geophysical variables at the different spatial scales specified above, we applied random forest analysis (hereafter, RF; Breiman 2001). RF is a recently developed, machine-learning technique that may be used for predictive modelling of multiple outputs from large input datasets. We executed this analysis using the R package ‘extendedForest’ (Smith et al. 2011), which is an updated version of the R package ‘randomForest’ and reduces bias when predictors are correlated. RF is a non-parametric technique derived from classification and regression trees (CART) (Breiman et al. 1984, Prasad et al. 2006). It uses an ensemble of decision trees with binary divisions, each capable of producing an output when presented with a set of input values (Cutler et al. 2007). Each tree is generated by bootstrap samples, leaving about a third of the overall sample for validation (the out-of-bag predictions – OOB). For regression modelling problems, such as the case we considered in this study, the tree response is the average of the results of all the trees (Breiman 2001, Cutler et al. 2007). As it uses the OOB samples (which are observations independent from those used to grow the trees), to calculate error rates and variable importance, no test data or cross-validation is required. However, the individual trees cannot be examined separately (Prasad et al. 2006) and RF calculates neither regression coefficients nor confidence intervals (Cutler et al. 2007), but instead provides a percentage of variance explained (%var) for each input independent variable. In addition, it is able to analyse large numbers of potential independent variables, and incorporates a method for calculating the importance of each independent variable in the final model that it provides (%imp; Smith et al. 2011). Previous analyses linking socioeconomic and environmental data to forest fragmentation or LUCC at the regional scale in other locations (e.g. Butler et al. 2004) and at the sub-regional level (e.g. Tyrell et al. 2004) have applied multiple linear regression using only one component of fragmentation (e.g., perimeter to area ratio) or LUCC change (e.g., forest loss). We found only one study (Bonilla-Moheno et al. 2012) that applied RF analysis to address similar relationships and investigated its application as different spatial scales, however it only considered changes in land cover types, and not in landscape fragmentation metrics. In the study reported here, we applied RF to one metric related to deforestation (DEFOR) and four to forest fragmentation (AI, CONNECT, ENN, and PD) as our dependent variables, linking them to socio-economic and bio-geophysical factors at the three spatial scales described above. Our data contained numerous cases in which variables are non-normally distributed, making this novel, non-parametric modelling approach a necessity.

3.3 Results

3.3.1 Deforestation and forest fragmentation quantification

The Atlantic Forest in Minas Gerais is estimated to have covered an area of 24,047,660 ha at the time of Europeans' first arrival in Brazil in the early 16th century (Galindo-Leal and Câmara 2003, Ribeiro et al. 2009). In 2003, less than 20% (19.19% = 4,615,811 ha) of this area of forest remained in the state. This is spread across the 19,904,945 ha covered by the 518 municipalities considered in this study. Deforestation data for the period that we studied, 2003-2011, are presented in Table 3.4 and Figure 3.3.

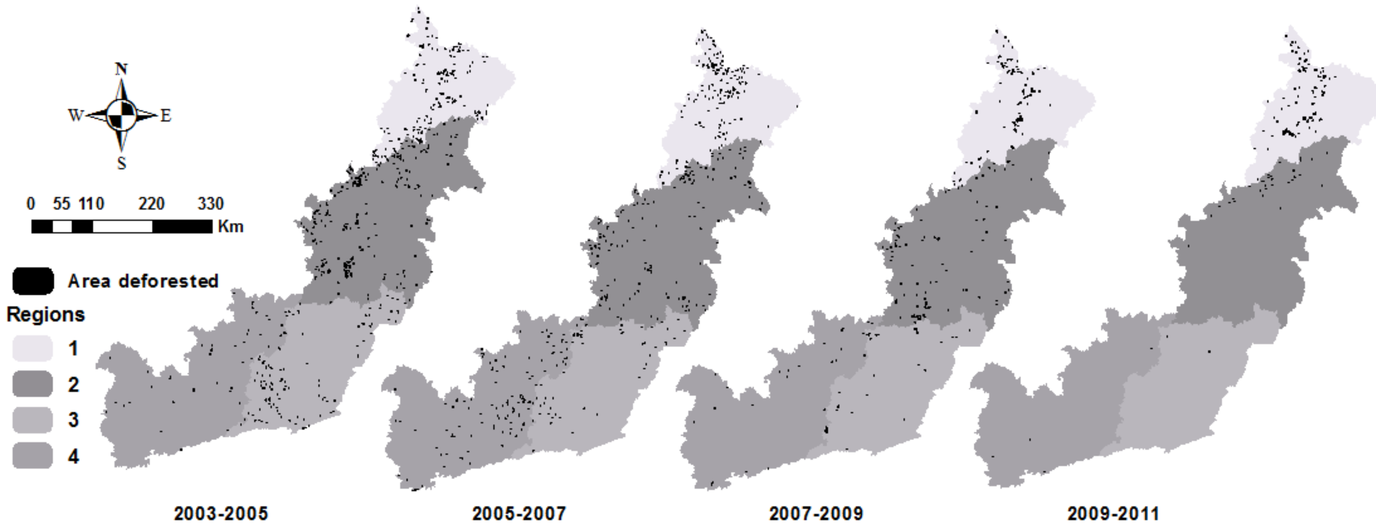


Figure 3.3 Area deforested in different periods of time across regions.

Table 3.4 shows the area deforested every two years between 2003 and 2011 in each of the four regions defined in Fig. 1B. The rate of deforestation generally (but not monotonically) decreased over this time period in Regions 2, 3 and 4, but increased in Region 1 (the northeastern-most part of the study area). Overall, there was a decrease in total deforested area, from 25,226 ha in 2003-2005 to 13,436 ha in 2009-2011 (a drop of 46.76%), but this difference fluctuated between positive and negative values in the intervening periods. In total, 77,968 ha of Atlantic Forest were deforested during the study period (4.538.843 ha of remaining forest area in 2011). This figure represents 1.69% of the remaining forest area in 2003 (4,615,811 ha).

Region 1 also experienced more deforestation than the other regions, in terms of both raw area values, and percentages of 2003 forested area (1.25% for Region 1 compared to 0.12–0.33% for Regions 2, 3 and 4). There is a monotonic decrease in both raw area and percentage of 2003 values of total deforestation from Region 1 to Region 4 (northeast to southwest).

In terms of landscape metrics patterns (Table 3.5), PD and AI decreased from Region 1 to Region 4, with the exception of Region 3, which had the smallest value of PD. Conversely, CONNECT and ENN showed no patterns in their north-south distribution, since the intermediate regions had the highest values of both metrics.

Table 3.4 Official deforested area figures obtained from the ecologic-economical zoning of Minas Gerais – ZEE-MG (Scolforo et al. 2008), including total deforested area (in hectares - ha and percentage - %) per year, regions and the entire Atlantic Forest (biome) in Minas Gerais.

Regions	**Region area (ha)**	**Total area deforested**									
		2003-2005		**2005-2007**		**2007-2009**		**2009-2011**		**2003-2011 (DEFOR)**	
		ha	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**	**ha**	**%**
1	3,258,197	9,814.50	0.0030	7,067.79	0.0022	11,018.61	0.0034	12,648.69	0.0039	40,549.59	0.0124
2	6,475,025	10,414.80	0.0016	5,112.81	0.0008	5,312.34	0.0008	679.95	0.0001	21,519.90	0.0033
3	4,341,869	3,226.23	0.0007	1,952.19	0.0004	2,727.99	0.0006	27.99	0.0000	7,934.40	0.0018
4	5,829,054	1,770.47	0.0003	3,632.31	0.0006	1,483.11	0.0003	78.93	0.0000	6,964.82	0.0012
Biome	19,904,146	25,226.00	0.0013	17,765.10	0.0009	20,542.05	0.0010	13,435.56	0.0007	76,968.71	0.0039

Table 3.5 Mean values of landscape metrics across regions and the entire Atlantic Forest (biome) in Minas Gerais.

Regions	**PD**	**AI (%)**	**CONNECT (%)**	**ENN (m)**
1	46.89	89.56	0.24	194.31
2	31.64	87.37	0.38	254.83
3	25.63	87.36	0.47	217.73
4	28.29	84.68	0.34	227.01
Biome	29.48	86.52	0.39	229.34

3.3.2 Deforestation and forest fragmentation metrics at multiple spatial scales Variance explained

The RF analysis provides strong evidence of patterns of co-variance between the independent variables (the socio-economic and bio-geophysical factors) and dependent variables (the deforestation and forest fragmentation metrics) at all of the spatial scales considered here (Table 3.6; Figure 3.4). Most of the models give high values of variance explained (up to 67.00% - see Figure 3.7).

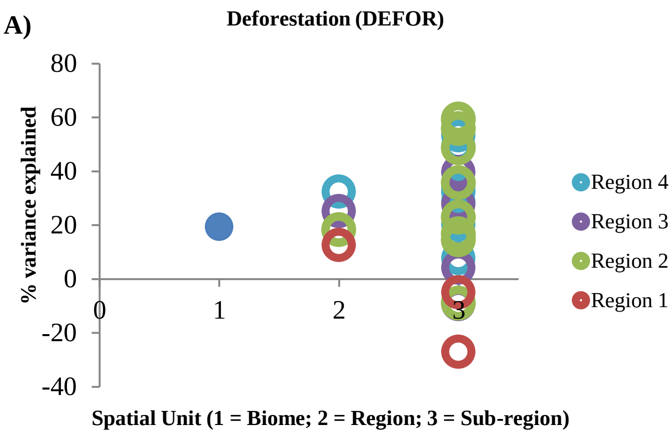
Table 3.6 Independent variable abundance for each dependent variable across the multiple spatial scales: all scales; regions and biome combined; and sub-regions.

	All Scales		**Regions/Biome**		**Sub-regions**	
	Ranking	**Count**	**Ranking**	**Count**	**Ranking**	**Count**
AI	Point_X	20	MinDist_nr	5	Point_X	15
	Point_Y	19	MinDist_re	5	Point_Y	14
	MinDist_re	17	MinDist_sm	5	Slo_mean	14
	MinDist_sm	17	Point_X	5	Annu_crops	12
	Slo_mean	17	Point_Y	5	MinDist_re	12
					MinDist_sm	12
					Rural_farms	12
CONNECT	Roads_dens	24	MinDist_ri	5	Roads_dens	19
	MinDist_ro	18	MinDist_ro	5	MinDist_ro	13
	MinDist_ri	17	Point_Y	5	MinDist_ri	12
	Alt_mean	15	Roads_dens	5	Alt_mean	11
	Point_Y	15			Slo_mean_d	11

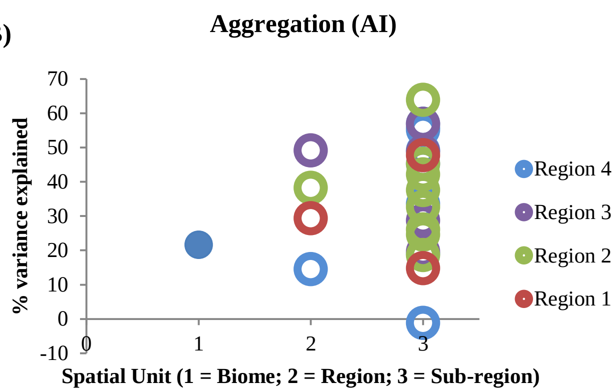
"Table 3.6, conclusion."

	All Scales		Regions/Biome		Sub-regions	
	Ranking	**Count**	**Ranking**	**Count**	**Ranking**	**Count**
DEFOR	Perm_crops	18	For_crops	5	Roads_dens	14
	Slo_mean_d	14	MinDist_nr	5	Slo_mean_d	11
	Min_comp	13	MinDist_sm	5	Alt_mean	9
	Den_pop	12	Point_X	5	Point_Y	9
	VA_agri04	12	Roads_dens	5		
	Unemp	12				
ENN	Alt_mean	22	MinDist_nr	5	Alt_mean	18
	MinDist_re	20	MinDist_re	5	Point_Y	16
	Point_Y	20	Perm_crops	5	Slo_mean	16
	Slo_mean	20	Point_X	5	MinDist_re	15
	MinDist_sm	19	Alt_mean	4	MinDist_sm	15
	Point_X	19	Income	4		
PD	Roads_dens	27	Roads_dens	5	Roads_dens	22
	Income	17	Alt_mean	4	Income	13
	Alt_mean	15	Income	4	Alt_mean	11
	MinDist_ri	15	MinDist_ri	4	MinDist_ri	11
			MinDist_sm	4		
			Perm_crops	4		

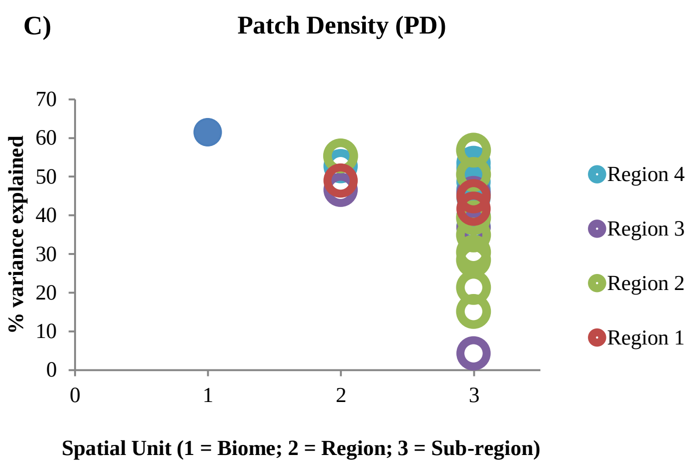
A)

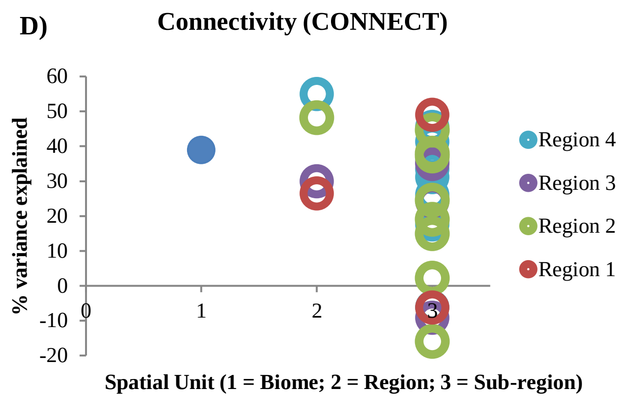

Deforestation (DEFOR)
% variance explained
80
60
40
20
0
-20
-40
0
1
2
3
Spatial Unit (1 = Biome; 2 = Region; 3 = Sub-region)
Region 4
Region 3
Region 2
Region 1

B)


Aggregation (AI)
% variance explained
70
60
50
40
30
20
10
0
-10
0
1
2
3
Spatial Unit (1 = Biome; 2 = Region; 3 = Sub-region)
Region 4
Region 3
Region 2
Region 1

C)


Patch Density (PD)
% variance explained
70
60
50
40
30
20
10
0
0
1
2
3
Spatial Unit (1 = Biome; 2 = Region; 3 = Sub-region)
Region 4
Region 3
Region 2
Region 1



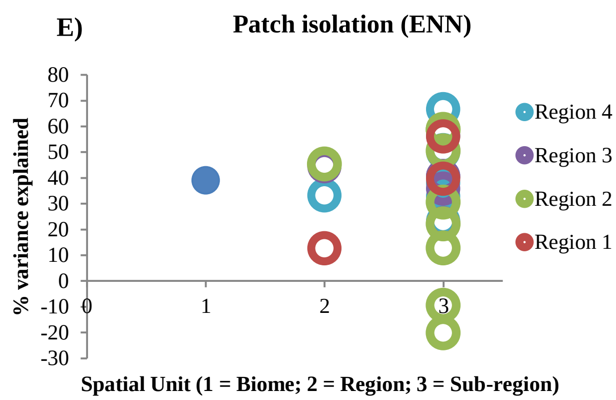


Figure 3.4 Ranking of aggregation units (sub-regions, regions, and biome) according to the percentage of variance explained considering all sub-regions, regions and the whole biome for the five dependent variables studied: AI - aggregation index; CONNECT - connectance index; DEFOR - the total amount of area deforested; ENN - mean Euclidean nearest-neighbour distance; PD - patch density.

The percentage of the variance of the deforestation and fragmentation metrics explained by our models varied across spatial scales, as well as between the metrics. For the total amount of area deforested (DEFOR), the mean value of variance explained stays approximately constant across all spatial scales, but the range of values spreads out as the spatial scale becomes finer. As a result, the models with both the highest and lowest percentages of variance explained occur at the sub-regional scale.

Patch density (PD) has arguably the clearest pattern of change in percentage variance explained with spatial scale, in that the highest values of variance explained occur for this variable at the largest (biome) scale, and decrease as the spatial scale becomes finer in a fairly consistent manner both in general and within each region.

Regarding the aggregation index (AI), there is, overall, a tendency for variance explained to increase as the scale becomes finer. For the connectance index (CONNECT), the percentage of variance explained is approximately the same at the biome scale as the average value across the four regions, but is lower on average at the sub-regional scale. Thus, the pattern here is somewhat similar to that for PD.

Finally, the mean Euclidean nearest-neighbour distance (ENN) presented a similar pattern to DEFOR, with the mean amount of variance explained being approximately the same across the three spatial scales, while the spread of values increases as one moves to finer spatial scales.

Overall, therefore, the general pattern is that for PD and CONNECT, the highest percentages of variance explained occur at the larger spatial scales, whereas for DEFOR, AI and ENN there is less of a monotonic trend with spatial scale, with both the highest and lowest percentages of variance explained occurring at the sub-regional scale.

3.3.3 Main factors and factor abundance

The relationships between deforestation/fragmentation metrics and socio-economic/bio-geophysical factors identified by RF analysis showed no single factor to be prominent in the models for all metrics at all spatial scales.

Nevertheless, a small number of factors appeared, with varying degrees of importance, across many of these models. Road density was among the most important factors in several models, and latitude, longitude, mean altitude, mean slope, and the minimum distance to the nearest reservoirs and steel mills were also common factors (Table 3.6).

Furthermore, we can identify patterns when we look at the most important factor affecting each metric individually at different spatial scales (Table 3.7). This is especially the case for the patch density (PD), for which the road density is the main factor in the large majority of models at all scales. Similarly, the road density was the most important variable in many of the models for the connectance index (CONNECT), especially at the larger scales (the biome-scale model and three of the four region-scale models), but also in several sub-regional models.

There was less consistency in terms of the most important factor across the models for the other deforestation and fragmentation metrics. Each of the models at biome and regional scales for the aggregation index (AI) had different most-important factors. However, the latitude (Point_Y) and longitude (Point_X) of the centroid of each municipality appeared as the most important factor in several models for this metric at regional and sub-regional scales, especially the latitude. The deforestation metric (DEFOR) also had no robust pattern of the most important factor across scales, but again the road density was the most common most-important-factor for this metric. Other factors that were also important in the DEFOR models were: forest crops, mean slope of areas deforested and latitude. Latitude was the most important factor in the biome-scale DEFOR model. Finally, in the ENN models, the mean altitude was the most important factor at biome scale and in two of the four regional scale models, and was also important in many sub-regional scale models, along with mean slope, latitude, and the minimum distance to the nearest reservoir.

Table 3.7 The most important socio-economic or bio-geophysical variable in the models generated by random forest regression analyses (RF) for each deforestation and forest fragmentation metric at the whole-biome, regional and sub-regional scales. See Tables 2 and 3 for explanations of abbreviations.

Spatial scales	AI			CONNECT			DEFOR			ENN			PD		
	Variable	**%var**	**%imp**	**Variable**	**%var**	**%imp**	**Variable**	**%var**	**%imp**	**Variable**	**%var**	**%imp**	**Variable**	**%var**	**%imp**
Biome	MinDist_sm	21.66	43.72	Roads_dens	44.51	115.65	Point_Y	19.41	31.2	Alt_mean	38.92	97.91	Roads_dens	61.47	200.45
Region 1	MinDist_nr	29.29	33.9	Roads_dens	26.47	37.08	ICMS_ecos	12.56	41.37	MinDist_re	12.51	21.31	Roads_dens	48.92	78.87
Region 2	Point_X	38.18	96.98	MinDist_ro	48.19	59.65	Roads_dens	18.33	60.11	Alt_mean	45.24	68.34	Roads_dens	55.31	111.01
Region 3	Point_Y	49.08	110.02	Roads_dens	29.9	60.58	Roads_dens	25.22	28.91	Alt_mean	44.38	79.43	Roads_dens	46.52	113.15
Region 4	Slo_mean	14.51	34.34	Roads_dens	54.96	90.59	Min_conc	32.47	44.71	Slo_mean	33.05	77.51	Roads_dens	52.67	142.71
Sub-region 1	Rural_fams	55.1	27.96	Den_pop	33.39	34.51	Min_comp	20.18	23.03	Slo_mean	30.54	33.41	Roads_dens	45.51	37.69
Sub-region 10	Emp	24.55	32.19	Rail_dens	24.48	21.38	Law_enf	35.71	35.49	Den_pop	-9.62	12.22	Roads_dens	39.36	40.09
Sub-region 11	Point_Y	56.89	52.01	Slo_min	34.86	53.67	For_crops	4.18	19.45	Point_X	40.48	50.64	Roads_dens	39.23	46.94
Sub-region 12	MinDist_ri	19	21.91	GPM_04	17.58	27.15	Roads_dens	34.32	33.46	Emp	37.93	27.8	Roads_dens	48.59	45.99
Sub-region 13	Rural_fams	19.62	19.72	ICMS_ecos	-9.25	8.82	MinDist_ro	-9.38	16.98	GGR_GPM	22.27	44.24	Roads_dens	4.19	26.8
Sub-region 14	MinDist_re	14.76	17.07	Roads_dens	-6.32	15.67	MinDist_ri	-5.03	12.69	Soil	55.91	42.45	Min_conc	44.84	36.93
Sub-region 15	Point_Y	48.73	52.34	Roads_dens	35.39	31.15	Income	27.79	32.95	Point_X	40.48	50.64	Roads_dens	45.68	44.76
Sub-region 16	Emp	32.61	21.71	MinDist_sm	18.97	28.75	Roads_dens	14.52	31.45	MinDist_nr	58.63	27.31	MinDist_re	56.82	29.56
Sub-region 17	Perm_crops	45.19	73.05	Roads_dens	31.09	82.87	Point_Y	28.72	39.29	Alt_mean	50.57	67.92	Roads_dens	52.13	101.58
Sub-region 18	Point_Y	33.58	35.08	Roads_dens	26.08	49.85	Slo_mean_d	53.26	34.44	Point_X	56.16	67.16	Roads_dens	44.53	58.69
Sub-region 19	Point_X	56.85	48.25	Roads_dens	41.26	36.7	For_crops	7.73	31.08	Point_X	49.9	47.98	Roads_dens	37.01	59.95
Sub-region 2	Alt_mean	26.07	20.17	Point_X	-16.02	8.04	Emp	55.79	19.19	Income	50.37	4.78	Soil	28.37	3.37
Sub-region 20	MinDist_sm	-1.2	13.06	Roads_dens	45.47	40.63	VA_agri04	49.37	43.12	Slo_mean	32.74	35.44	Roads_dens	53.49	57.38
Sub-region 21	Point_Y	42.16	34.57	MinDist_sm	44.57	37.93	Rural_fams	22.91	31.07	Annu_crops	-20.28	19.03	Roads_dens	15.05	23.89

"Table 3.7, concluison."

Spatial scales	AI			CONNECT			DEFOR			ENN			PD		
	Variable	**%var**	**%imp**	**Variable**	**%var**	**%imp**	**Variable**	**%var**	**%imp**	**Variable**	**%var**	**%imp**	**Variable**	**%var**	**%imp**
Sub-region 22	Annu_crops	28.72	46.19	Roads_dens	24.18	29.8	Slo_mean_d	39.68	60.22	Alt_mean	35.71	42.68	Roads_dens	36.72	25.17
Sub-region 3	Point_Y	18.55	23.69	Roads_dens	37.97	56.1	For_crops	-8.98	23.14	Point_Y	58.55	62.04	Roads_dens	34.86	55.46
Sub-region 4	Point_Y	49.14	31.17	Law_enf	36.79	42.63	Slo_mean_d	32.17	35.84	MinDist_re	23.08	31.35	Roads_dens	46.62	37.09
Sub-region 5	MinDist_sm	55.84	37.14	Unemp	-6.18	19.47	Rural_fams	33.01	24.86	Slo_mean	66.53	42.73	Rural_fams	39.8	34.28
Sub-region 6	MinDist_re	37.54	34.11	MinDist_ro	37.11	34.48	Point_Y	48.69	43.46	Perm_crops	21.99	36.16	Income	50.51	39.55
Sub-region 7	Point_X	63.9	55.12	GGR_GPM	14.87	19.59	MinDist_ro	16.77	23.52	MinDist_re	12.64	33.97	Income	21.3	26.34
Sub-region 8	GPM_04	45.16	33.84	Roads_dens	2.12	17.23	GPM_04	59.42	38.98	Alt_mean	30.4	32.65	GGR_GPM	30.37	29.14
Sub-region 9	Point_Y	47.69	32.98	Point_Y	49.01	30.83	Roads_dens	-27.09	24.56	MinDist_re	39.58	33.94	Roads_dens	41.66	46.54

We also quantified how often each factor appeared in the models for each metric (Table 3.6). In this sense, the PD models once again presented the strongest pattern, in that the road density appeared in every model across all spatial scales. The DEFOR and CONNECT models showed a pattern similar to the PD models, in that the road density was also the most common factor in these models across all spatial scales. Other factors common in the DEFOR models were forest crops, the minimum distances to the nearest natural reserves and steel mills, and longitude at larger spatial scales, and mean slope of areas deforested, mean altitude and latitude at the sub-regional scale. In the CONNECT models, the minimum distance to the nearest river was also a common factor.

The models of the other metrics were also similar in terms of the most common factors. The most common factors in the AI models were the minimum distances to the nearest reservoirs and steel mills, latitude and longitude. The minimum distance to the nearest protected area was also abundant in the models at biome and regional scales while mean slope, the amount of annual crops and the number of rural family farms were common factors in the sub-regional scale models. Finally, the most common factor in the ENN models was the mean altitude. The minimum distance to the nearest reservoir was also a common factor across ENN models at all spatial scales. The minimum distance to the nearest natural reserve, the amount of permanent crops, and longitude were more common in the biome and regional-scale models, while latitude and mean slope were the most common factors in the sub-regional scale ENN models.

3.4 Discussion

3.4.1 Deforestation and forest fragmentation quantification

Our study shows the ongoing occurrence of deforestation processes in the Atlantic Forest of Minas Gerais. The total area deforested during the period investigated (2003-2011) represents 1.69% of the current remaining forest area. While this could be thought as a small percentage it is, nevertheless, alarming, since this biome is considered a hotspot of biodiversity (Mittermeier et al. 2005, Zachos and Habel 2011), and there is specific legislation prohibiting its deforestation (Law n.11.428; Brazil 2006). According to the Atlantic Forest Law (Brazil 2006), exploitation and suppression of Atlantic forest remnants are permissible only by decision of the environmental agencies, such as the Brazilian Institute of Environment and Renewable Natural Resources – IBAMA or the State Department of Environment and Sustainable Development (SEMAD) at the state level. In addition, the Forest State Institute of Minas Gerais – IEF (2013), which is linked to SEMAD, has identified illegal deforestation within the state's borders, and most deforested areas appear to correspond to illegal actions. It was not possible, however, for us to identify which of the deforestation processes were illegal.

We observed an overall declining trend in the rate of deforestation during the period under study (Table 3.4; Figure 3.3), possibly because of more intense surveillance resulting from the Atlantic Forest Law (Brazil 2006). There was a trend of deforestation being higher in the North of the state (Region 4), and declining southwards through Regions 3, 2 and 1. This trend becomes stronger over the time period for which data is presented in Table 3.4. This trend is also evident when the data is presented in terms of the percentage of remaining forest area deforested in each 2-year period, but it may still reflect a correlation between the area deforested and the amount of remaining forest area, since the latter is also higher towards Northern (Carvalho and Scolforo – unpublished data), in that law enforcement may be more strictly applied where there is less remaining forest.

The spatial structure of the remaining forest showed a northwards trend of increasing patch density (PD). We observed trends in the same direction for forest aggregation (AI). However, there was no clear pattern of values distribution for forest connectance (CONNECT) and isolation (ENN).

3.4.2 Deforestation and forest fragmentation associated factors at multiple spatial scales

This study is the first multi-scale assessment of relationships between deforestation and forest fragmentation metrics and socio-economic and bio-geophysical factors in the Brazilian Atlantic Forest. To our knowledge, this is also the first time that Random Forest (RF) analysis has been used to identify relationships between these types of metrics and factors, using a multi-scale approach (but see Bonilla-Moheno et al. 2012 for a previous two-scale RF study of land cover changes). We discuss our findings, firstly, by comparing the variance explained by the models produced by our analysis across spatial scales and between metrics, and, subsequently, by interpreting the implications of the most common and important factors found by these models.

The deforestation and fragmentation metrics follow different patterns at different spatial scales in terms of the amount of their variance that the RF output models explain. There is a tendency for some metrics to be best explained in some of the sub-regions, while others are better explained at larger spatial scales. Deforestation (DEFOR), forest patch isolation (ENN), and forest aggregation (AI) had the highest percentages of variance explained in some of their sub-regions.

Forest patch connectance (CONNECT) and patch density (PD) had higher percentages of variance explained at the biome and regional scales. This variation in metrics variance explanation across scales highlights the presence of the modifiable areal unit scale problem (MAUP; Openshaw and Taylor, 1979) in our modelling. This consists of a "variation in results that may be obtained when the same areal data are combined into sets of increasingly larger areal units of analysis" (Openshaw and Taylor, 1979). In our investigation, we were able to address this issue, identifying the metrics that can be affected by the MAUP scale problem. While the MAUP reflects the nature of hierarchically-structured systems in the real world, there is no real impediment to understanding spatial phenomena if MAUP is adequately recognized and dealt with explicitly (Jelinski and Wu 1996). We reinforce, then, that the MAUP must be investigated, since we have demonstrated here that different relationships between the factors related to deforestation and forest fragmentation may arise at different spatial scales.We used the same spatial unit (municipalities) across the scales.

Different arguments can be made as to why the deforestation metric (DEFOR) is best explained in some sub-regions in the Atlantic Forest of Minas Gerais. Firstly, this biome covers a huge area, so a large degree of difference in practices across the entire biome might make for an unclear set of relationships at the biome and regional scales (Ribeiro et al. 2011). A more spatially-varied pattern of deforestation and of its relationship to socio-economic and bio-geophysical factors across the biome, and from sub-region to sub-region may be caused by a combination of factors: the long-term background of degradation and deforestation in the Atlantic Forest, and the resultant unevenness of its remaining area, together with regional differences in the way that controls on deforestation have been enforced. In specific sub-regions that have a larger deforested area, we observed a stronger connection between DEFOR and these factors than in sub-regions where the amount of deforestation was very low or even non-existent.

This may explain the wide variability in the percentage of variance explained across the sub-regions. Furthermore, we found that the main factor explaining deforestation at the biome scale was the geographical location of municipalities. This also reinforces the idea that the patterns and drivers of deforestation are highly spatially specific and therefore is best explained at finer spatial scales (Geist and Lambin 2002).

Forest aggregation (AI) and isolation (ENN) were also best explained in some of the sub-regions. These fragmentation measurements are negatively correlated with each other, since the higher the forest aggregation, the lower the patch isolation (McGarigal et al. 2012). Both followed trends similar to that observed for deforestation, since deforestation is also associated with where the remaining forest is, and consequently with forest aggregation. This is consistent with the idea that the deforestation that has been undertaken in the Atlantic Forest has resulted in a spatially-aggregated pattern of remaining forest. Thus, at larger spatial scales relationships between deforestation and forest fragmentation metrics and socio-economic and bio-geophysical variables are likely to be distorted by spatial variations in historical patterns of deforestation, whereby it was more intense in the south of the state than the north (Carvalho and Scolforo – unpublished data). At sub-regional scales, however, the historical pattern of deforestation is likely to be more homogeneous, implying that the relationships will have greater explanatory capability. On the other hand, forest patch density (PD) and connectance (CONNECT) had greater values of variance explained in the models at larger spatial scales, implying that these variables are less affected by local differences in the underlying historical pattern of deforestation.

Note that, in some models, a substantial part of the variance was not explained by the factors considered, even though we used a very large dataset covering a great variety of independent variables. Thus, there are clearly attributes of municipalities that remain uncaptured in our data that influence the nature of deforestation of the Atlantic Forest.

In summary, our results suggest that deforestation, forest aggregation and isolation have patterns that are less predictable from our set of socio-economic and bio-geophysical variables at the full biome and regional scales than they are at sub-regional scales in some areas of the Atlantic Forest in Minas Gerais. Thus, they can be thought of as having more coherent and comprehensible patterns of spatial variation at smaller spatial scales. Conversely, patch density and connectance present more predictable patterns at the biome scale than they do at sub-regional scales.

3.4.3 Main factors and factor abundance

In accordance with our expectations and the findings of other studies, we found that there is no single factor influencing all of the deforestation and fragmentation metrics, but a set of different factors affecting each one. Geist and Lambin (2002) obtained similar results when they investigated the causes of deforestation of tropical forests. They concluded that a combination of factors causes forest loss, and that these depend on historical and geographical context. Here, we find that this is the case for metrics that quantify the spatial structure of forest fragmentation due to deforestation as well as simple measures of the amount of deforestation. According to Seabloom et al. (2002), deforestation (and consequently forest fragmentation) is recognized worldwide as a process that follows non-random patterns. Endorsing the results of other studies, our study supports the theory that deforestation and fragmentation are influenced by a wide variety of factors in tropical regions (Geist and Lambin 2002, Bonilla-Moheno et al. 2012). These factors are found, to a certain extent, to be specific to the area and scale of the study. Laurance et al. (2001) and Gardner et al. (2009) found that soil fertility, economic interests, proximity to urban settlements and roads are among the important factors that drive deforestation and fragmentation in tropical regions. Other studies in the Atlantic Forest found that deforestation and regeneration processes are influenced by factors such as topography, land use, and the distribution of urban areas (Silva et al. 2007, Teixeira et al. 2009, Freitas et al. 2010). Our study partially agrees with these studies: we found that factors from different categories, mainly from Accessibility, Geographical location, Topography, and Distance categories influence deforestation and fragmentation most.

We identified specific factors influencing each metric, the importance of which varied across scales and between regions and sub-regions. Accessibility, specifically the road density, was the most important and abundant category affecting patch density, deforestation and forest patch connectance across all spatial scales. This finding is consistent with intuition, since roads serve as fragmenting features (Forman and Alexander 1998, Butler et al. 2004), subdividing forests, increasing the number of forest patches, and reducing forest connectance. Roads have few positive, neutral and numerous negative environmental impacts. Positive impacts include increasing accessibility (Leinbach 1995), which can also be negative since this facilitates deforestation (Laurance et al. 2001). Negative impacts include habitat loss, degradation, and fragmentation, direct wildlife mortality, and road avoidance behaviours by wildlife (Forman and Alexander 1998).

Beyond this identification of road density and accessibility being the most important element in our models, a complex and varying mix of factors from most of the categories in Table 3.3 are found to be important. The variables that are most often identified as being of importance in these models are from the Distance, Geographical, Topographic and Natural categories, with the Human and Institutional categories contributing to the models less commonly. This implies that it is the spatial distribution of the forest in relation to the underlying landscape structure i.e. the layout of the natural and man-made infrastructure (rivers, reservoirs, roads, steel mills etc.), the underlying topography and the distribution of agricultural and protected land, that is more important in determining the patterns of deforestation and forest fragmentation, and that human demographic, employment, income and institutional factors are less important in determining these patterns. This appears at first sight to conflict with the view mentioned in the introduction, that socio-economic processes are thought to be the primary factors associated with LUCC, and that these determine the structure, function and dynamics of most landscapes (Lorenzoni et al. 2000, Wang and Zhang 2001, Wu and Hobbs 2002). Our findings can be reconciled with this view by noting that it is the structure, function and dynamics of the landscape that is represented in the independent variables that our analysis has found to be important in models of deforestation and fragmentation, and that these will have been determined by past socio-economic activities. Thus, there appears to be an asynchronicity in the way in which socio-economic drivers are related to patterns of deforestation, in that their current nature (which is what is represented in our input dataset) is not strongly related to these patterns, but their past nature may well have been, albeit indirectly via the way in which it may have governed landscape development.

3.5 Conclusions

Our study has provided a multi-scale assessment showing the relationships between deforestation/forest fragmentation metrics and socio-economic/bio-geophysical factors in the Atlantic Forest, using Random Forests (RF) analysis and a very large dataset. We have shown that RF analysis can be used to identify links between these sorts of metrics and factors. We have found that the metrics behave differently at different spatial scales in terms of the amount of their variance that can be explained by the factors in our dataset, evidencing the presence of the modifiable areal unit problem across the spatial scales we investigated. We found deforestation and forest patch aggregation and isolation to be best explained at the finest scale we considered (sub-regional scale). We infer that this is due to their more coherent relationships with socio-economic and bio-geophysical factors at this relatively local scale than at the regional or biome scale, where historical and geographical differences may alter these relationships from place to place. Conversely, we found patch density and forest connectance to be best explained at larger scales, and infer that these metrics are more coherently related to socio-economic and bio-geophysical factors at these scales. We also found that there is no unique factor affecting all metrics across all spatial scales, but a set of factors from different categories. Nevertheless, road density was found to be the single most prevalent factors in the models generated by our RF analysis, and other common factors came from categories describing the spatial distribution of the natural, agricultural and infrastructural elements of the landscape. Factors describing patterns of population, employment and legal institutions were found to be less important in our models. Finally, we have also shown that deforestation processes are still occurring in the Atlantic Forest of Minas Gerais, despite specific legislation making this illegal.

Acknowledgements

The authors would like to thank the Federal University of Lavras (UFLA) for providing the data. L. Zanella would like to acknowledge support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), who provided a PhD scholarship, and Ricardo Solar and Teotônio Soares de Carvalho for assistance in the statistical analysis.

REFERENCES

Bonilla-Moheno M, Mitchell Aide T, Clark ML. 2012 The influence of socioeconomic, environmental and demographic factors on municipality scale land-cover change in Mexico Regional Environmental Change 12:543–557.

Brazil. 2006. Law N. 11428, of 22 of December, 2006 Regulation for use and protection of native vegetation of the Atlantic Forest biome, and other matters Available from: http://wwwplanaltogovbr/ ccivil_03 /_ato2004 - 2006/2006/lei/l11428htm (Accessed January 2015).

Breiman L, Friedman JH, Olshen RA, Stone CI. 1984. Classification and regression trees. Belmont Calif.: Wadsworth.

Breiman L. 2001. Random forests. Mach Learn 45:5–32.

Brooks TM, Mittermeier RA, da Fonseca GAB, Gerlach J, Hoffmann M, Lamoreux JF, Mittermeier CG, Pilgrim JD, Rodrígues ASL (2006) Global biodiversity conservation priorities. Science 313:58–61.

Butler BJ, Swenson JJ, Alig RJ. 2004. Forest fragmentation in the Pacific Northwest: quantification and correlation. Forest Ecol Manage 189:363–373.

Carrato JF. 1968. Igreja, Iluminismo e escolas mineiras coloniais. São Paulo: Nacional/Edusp.

Carvalho LMT, Scolforo JRS. 2008. Inventário Florestal de Minas Gerais: Monitoramento da Flora Nativa 2005-2007. Lavras, Editora: UFLA, 357p.

Carvalho LMT, Scolforo JRS. Unpublished data. Inventário Florestal de Minas Gerais: Monitoramento da Flora Nativa 2007-2009; 2009-2011.

Cutler DR, Edwards TC, Beard KH, et al. 2007. Random forests for classification in ecology Ecology 88:2783–92.

Deng X, Zhao C, Lin Y, Zhang T, Qu Y, Zhang F, Wang Z, Wu F. 2014. Downscaling the Impacts of Large-Scale LUCC on Surface Temperature along with IPCC RCPs: A Global Perspective. Energies. 7(4):2720-2739.

FAO –Food and Agriculture Organization of the United Nations. 2008. UN collaborative programme on reducing emissions from deforestation and forest degradation in developing countries (UN-REDD). Rome, 2008.

Forman RTT, Alexander LE. 1998. Roads and their major ecological effects Annual Review of Ecology and Systematics 29: 207–31.

Freitas SR, Hawbaker TJ, Metzger JP. 2010. Effects of roads, topography, and land use on forest cover dynamics in the Brazilian Atlantic Forest Forest Ecology and Management 259:410-417.

Galindo-Leal C, Camara I. 2003. The Atlantic Forest of South America: Biodiversity Status, Threats, and Outlook, Center for Applied Biodiversity Science at Conservation International, Island Press, Washington.

García-Barrios L, Galván-Miyoshi YM, Valdivieso-Pérez IA, Masera OR, Bocco G, Vandermeer J. 2009. Neotropical forest conservation, agricultural intensification, and rural out-migration: the Mexican experience Bioscience 59:863–873.

Gardner TA, Barlow J, Chazdon R, Ewers RM, Harvey CA, Peres CA, Sodhi NS. 2009. Prospects for tropical forest biodiversity in a human-modified world Ecol Lett 12:561–582.

Geist HJ, Lambin EF. 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. BioScience. 52 (2): 143-150.

Heistermann M, Müller C, Ronneberger K. 2006. Land in sight? Achievements, deficits and potentials of continental to global scale land-use modeling Agric Ecosyst Environ 114:141-158.

IBGE – Brazilian Institute of Geography and Statistics. 2015. Estados. Minas Gerais. Available from: http://www.ibge.gov.br/estadosat/perfil.php?sigla=mg. (Accessed May 2015).

IBGE. 2004. Mapa de Biomas e de Vegetação Available from ftp://ftpibgegovbr/Cartas_e_Mapas/Mapas_Murais/ (Accessed August 2014).

IEF– Forest State Institute of Minas Gerais. 2013. Fiscalização combate desmatamento da Mata Atlântica em Minas. Available from: http://www.ief.mg.gov.br/noticias/1/1687-fiscalizacao-combate-desmatamento-da-mata-atlantica-em-minas- (Accessed March 2015).

Jelinski DE, Wu J. 1996. The modifiable areal unit problem and implications for landscape ecology. Landscape Ecology 11(3):129-140.

Lambin EF, Geist HJ. 2006. Land-Use and Land-Cover Change: Local Processes and Global Impacts – The IGBP Ser. Berlin: Springer-Verlag.

Lambin EF, Geist HJ, Lepers E. 2003. Dynamics of land-use and land-cover change in tropical regions Annu Rev Environ Resour 28:205–241.

Lambin EF, Turner BL II, Geist H, Agboola S, Angelsen A, Bruce JW. 2001. The causes of land-use and land-cover change: moving beyond the myths Global Environmental Change 11:261–269.

Laurance WF, Cochrane MA, Bergen S, Fearnside PM, Delamônica P, Barber C, D'angelo SE, Fernandes T. 2001. The Future of the Brazilian Amazon Science 291:438-439.

Leinbach T. 1995. Transport and third world development: review, issues, and prescription Transportation Research A 29(5):337-344.

Lira PK, Tambosi LR, Ewers RM, Metzger JP. 2012. Land-use and land-cover change in Atlantic Forest landscapes For Ecol Manage 278:80–89.

Lorenzoni I, Jordan A, Hulme M, et al. 2000. A co-evolutionary approach to climate change impact assessment: Part I Integrating socio-economic and climate change scenarios Glob Environ Chang 10:57–68.

McGarigal K, Cushman SA, Ene E. 2012. FragStats v4: Spatial Pattern Analysis Program for Categorical and Continuous Maps Computer software program produced by the authors at the University of Massachusetts, Amherst Available from: http://wwwumassedu/landeco/research/fragstats/fragstatshtml (Accessed May 2013).

Mittermeier R, Turner W, Larsen F, Brooks T, Gascon C. 2011. Global Biodiversity Conservation: The Critical Role of Hotspots. Berlin: Springer Berlin Heidelberg.

Mittermeier RA, Gil RP, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, Lamoreux J, Fonseca GAB. 2005. Hotspots revisited: earth's biologically richest and most endangered terrestrial ecoregions. 2nd ed. Boston: University of Chicago Press.

Musaoglu N, Tanik A, Kocabas V. 2005. Identification of land-cover changes through image processing and associated impacts on water reservoir conditions. Environ Manage 35(2):220–230.

Openshaw S, Taylor PJ.1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. In: Wrigley N. (ed). Statistical applications in the spatial science. London: Routledge. p. 127-144.

Prasad AM, Iverson LR, Liaw A. 2006. Newer Tree Classification and Techniques: Forests Random Prediction Bagging for Ecological Regression Ecosystems 9:181–199.

Ribeiro MC et al. 2009. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation Biological conservation, 142:1141-1153.

Ribeiro MC, Martensen AC, Metzger JP, Tabarelli M, Scarano F, Fortin M-J. 2011. The Brazilian Atlantic Forest: a shrinking biodiversity hotspot In: Zachos FE, Habel JC (Org) Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas London New York: Springer-Verlag Berlin Heidelberg. 1:405-434.

Riitters KH. et al. 1995. A factor analysis of landscape pattern and structure metrics. Landsc Ecol 10:23–39.

Scolforo JRS, Oliveira AD de, Carvalho LMT. 2008. Zoneamento ecológico-econômico do estado de minas gerais: Componente socioeconômico. Lavras: Editora da UFLA.

Scolforo JRS, Carvalho LMT. 2006. Mapeamento e inventário da flora nativa e dos reflorestamentos em Minas Gerais Lavras: Editora da UFLA 288p.

Scrieciu SS. 2007. Can economic causes of tropical deforestation be identified at a global level? Ecol Econ 62:603–612.

Seabloom EW, Dobson AP, Stoms DM. 2002. Extinction rates under nonrandom patterns of habitat loss. PNAS. 99(17): 11229-1123.

Silva WG, Metzger JP, Simões S, Simonetti C. 2007. Relief influence on the spatial distribution of the Atlantic Forest cover on the Ibiúna Plateau, SP. Braz J Biol 67(3):403-411.

Smith SJ, Ellis N, Pitcher CR. 2011. Conditional variable importance in R package extendedForest Available from http://gradientforestr-forger-projectorg/Conditional-importancepdf (Accessed January 2014).

SOS Mata Atlântica and INPE - National Institute for Space Research. 2008 Atlas dos remanescentes da Mata Atlântica 2000-2005 Fundação SOS Mata Atlântica, São Paulo.

Teixeira AMG, Soares-Filho BS, Freitas SR, Metzger JP. 2009. Modeling landscape dynamics in an Atlantic Rainforest region: Implications for conservation. For Ecol Manage 257:1219–1230.

Turner II BL, Skole D, Sanderson S, Fischer G, Fresco L, Leemans R. 1995. Land-use and land-cover change, Science/research plan. IGBP report Nº 35. HDP Report Nº 7. 132 p.

Wang Y, Zhang X. 2001. A dynamic modeling approach to simulating socioeconomic effects on landscape changes. Ecol Modell 140:141-162.

Wu J, Hobbs R. 2002. Key issues and research priorities in landscape ecology: an idiosyncratic synthesis. Landscape Ecology 17: 355–365.

Zachos FE, Habel JC. 2011. Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas. Berlin Heidelberg: London New York Springer-Verlag.

CHAPTER 4

SPECIES DISTRIBUTION MODELLING DEMONSTRATES THE NEED FOR EXPANSION OF PROTECTED AREAS IN BIODIVERSITY HOTSPOTS OF MINAS GERAIS, BRAZIL

Publication status: *In prep.* for submission to *Biological Conservation*

ABSTRACT

We combined potential distribution data of threatened plant species with remaining forest fragments and protected areas (PAs) in Minas Gerais, Brazil, in order to assess the effectiveness and representativeness of PAs for threatened species. We performed potential distribution models for eight red-listed plant species, according to the International Union for Conservation of Nature (IUCN). Data from the Vegetation Monitoring System held in Minas Gerais and complementarily from SpeciesLink and NeoTropTree databases were used to model the potential distribution of the species, based on the Maximum Entropy (MaxEnt) method. We combined maps showing areas with > 50% of environmental suitability for the occurrence of the studied species and the maps of natural vegetation remnants and the PAs. Areas with the highest potential occurrence of threatened species that did not overlay with PAs were considered as gaps in protection, and therefore, priority areas for conservation. A variety of environmental variables explained habitat suitability, according to the species considered. The map of environmental suitability for all species combined showed significant gaps in the network of PAs. We recognized at least three sites of high environmental suitability for most of the species we studied: one in the Southeast, one Central and another one in the North of Minas Gerais. The total area under protection in Minas Gerais represents less than 7% of its territory. The total potentially suitable area for at least one of the threatened species we considered that within the 246 PAs corresponds to 4.61% of the total area of the state. Therefore, the amount of land under protection is considered to be far from sufficient for adequate conservation of the species. The main gaps of protection are in the Atlantic Forest biome, which also has the highest number of PAs. We believe that this result is related to the geographic distribution of species studied, as they all occur in the Atlantic Forest. The Atlantic Forest has also been considered as an object of study in several studies, so it is possible that many species from the Cerrado and Caatinga biomes are missing from the red lists due to lack of studies. As these two biomes have only a few PAs in Minas Gerais, they are also a priority for conducting inventories and creating natural reserves.

Keywords: Brazilian Atlantic Forest. Cerrado. Caatinga. Tropical environments. Maximum entropy modeling. Minas Gerais State. Potential distribution. Species Distribution Modeling – SDM. Environmental suitability.

4.1 Introduction

Impacts of human activities upon biodiversity are increasing at unprecedented rates, resulting in widespread species extinctions (Dirzo et al. 2014, Pimm et al. 2014). The great majority of species respond negatively to human disturbances triggering rapid changes to ecosystems worldwide (Baillie et al. 2004, Fahrig and Rytwinski 2009). Habitat destruction, and associated degradation and fragmentation, are considered the main threats faced by most species (Baillie et al. 2004, Brooks et al. 2006).

Several strategies for nature conservation have emerged in response to the biodiversity crisis. The establishment of Protected Areas (hereafter, PAs) is a key strategy (Butchart et al. 2010), and their proposition must follows systematic conservation planning carried out in several stages (Margules and Pressey 2000). The identification of species under threat from extinction (e.g. Biodiversitas 2005, MMA 2008, IUCN 2014), and the identification of regions of mega-diversity (in terms of species richness, endemism and/or genetics) which are under the greatest threat (e.g. Mittermeier et al. 2005, Mittermeier et al. 2011, Zachos and Habel 2011) are examples of the issues that should be assessed before proposing a new PA, issues that models of species' potential distributions can help to overcome.

Species distribution models (hereafter, SDM) are considered important tools for identifying priority areas to be protected (Ferrier et al. 2002). These models are based on species' environmental requirements, and assume that the higher the suitability of an environment is for a species, the more persistent populations of that species will be within that environment, at least in the short term (Araújo et al. 2002). They have been found to provide information that is essential to guiding conservation and recovery efforts (Elith et al. 2006, Phillips et al. 2006). The application of such models to threatened species has contributed strongly to improving the effectiveness of more protected area networks (Gaston

1996). However, occurrence data on threatened species is sparse (Ferrier et al. 2002, Engler et al. 2004), making it difficult to model their potential distribution.

SDM's have become increasingly popular tools for predicting the potential distributions of species. SDM's have been used for a variety of purposes, such as: improving detection of new areas of occurrence of species by guiding field surveys (Bourg et al. 2005), assessing possible impacts arising from climate change (Thuiller et al. 2005a), estimating risks of biological invasions (Thuiller et al. 2005b) and supporting conservation decisions (Ferrier et al. 2002, Lemes et al. 2014, Ferro et al. 2014). These models relate data on the occurrence of species with environmental layers from the area of interest. Several alternative methods of modelling species distributions have been proposed (Guisan and Zimmermann 2000, Guisan and Thuiller 2005, Elith et al. 2006, Wisz et al. 2008). The outcomes of these models have shown great potential across a large range of applications. However, only a few modelling efforts have been applied to threatened species (Engler et al. 2004) or to assess the extent to which existing protected areas support their conservation.

The state of Minas Gerais, Brazil, host three important biomes: Atlantic Forest, Cerrado and Caatinga (IBGE 2004), the first two of which are considered to be hotspots of biodiversity (Mittermeier et al. 2005, Mittermeier et al. 2011, Zachos and Habel 2011) because of their importance in terms of diversity, and because they are under serious threat. The third biome, the Caatinga, is also very diverse and is currently under threat (Castelletti et al. 2004). Threats to all three biomes include many anthropogenic factors, such as deforestation and forest fragmentation, fire, and expansion of agriculture and pasture (Mittermeier 2005, Mittermeier et al. 2011, Zachos and Habel 2011). As a result, Minas Gerais has 400 threatened species within its boundaries (IUCN 2014). The geographical distribution of most of these species remains unknown. Application of SDM could provide essential information for an effective conservation planning for all of these biomes. Therefore, the aim of this study was to assess the effectiveness of existing protected areas and the conservation status of threatened plant species in the Atlantic forest, Cerrado and Caatinga Biomes of Minas Gerais state, Brazil. Our specific objectives were: (1) to identify the potential distribution of threatened plant species in Minas Gerais, based on measures of environmental suitability; (2) to combine maps of these potential distributions with maps of remaining natural vegetation fragments and PAs, to assess gaps in the protection of the species studied; and (3) to recommend priority areas for conservation based on the identification of environmentally suitable areas for the species considered in this study.

4.2 Methods

4.2.1 Study area

The state of Minas Gerais is located in Southeastern Brazil between latitudes 14º 03' 28'' S and 23º 07' 02'' S and longitudes 51º 07' 02'' W and 39º 49' 58'' W. It covers an area of 58,652,212 ha and is split into 853 municipalities, ranging in area from 285 ha to 1,071,696 ha. It has three biomes within its limits: Cerrado, Caatinga and Atlantic Forest (IBGE 2004). The Cerrado and Atlantic Forest biomes are considered to be hotspots of biodiversity in Brazil (Mittermeier et al. 2005, Mittermeier et al. 2011, Zachos and Habel 2011) (Figure 4.1).

Figure 4.1 Study area location. The inset maps on the left show the location of Brazil in the South America in the upper map, and the Minas Gerais State within Brazil in the lower map.

4.2.2 Species selection and occurrence data

We selected all plant species red-listed by the IUCN (2014) in Minas Gerais that were also present in the lists of the Vegetation Monitoring System Project (Scolforo and Carvalho 2006). This project sampled 169 fragments of remaining natural vegetation across the state, where plants with diameters equal to or greater than 5 cm at human chest height were sampled in more than 4.400

parcels. In this study, 21 higher plant species were selected (Table 4.1). The locations of occurrence of each of these species were compiled from three sources:

1. The inventory reports from the Vegetation Monitoring System Project (Scolforo and Carvalho 2006);
2. A search of the database of the national Herbaria network (SpeciesLink, available at http://splink.cria.org.br/);
3. A search of the Neo Trop Tree data base (http://prof.icb.ufmg.br/treeatlan/), which gathers data on biogeography, diversity and conservation of tree flora of the Neotropical Region.

From the initial list, we selected only the species with at least ten points of occurrence in Minas Gerais (Table 4.2, Figure 4.2 and 4.3). We used all records from the Vegetation Monitoring System Project and the NeoTropTree database. Species in the national Herbaria network database that did not have geographical information associated with them were disregarded. This led to selection for modelling of a sub-set of eight threatened species from the initial 21 species. Of these, one species (*Mimosa bimucronata* (DC.) Kuntze), occurs exclusively in the Atlantic Forest; three species (*Araucaria angustifolia* (Bertol.) Kuntze, *Pereskia aculeata* Miller and *Podocarpus lambertii* Klotzsch) occur in both the Atlantic Forest and the Cerrado biomes; and four species (*Andira fraxinifolia* (Benth.), *Cereus jamacaru* DC, *Pereskia grandifolia* Haw., and *Platypodium elegans* Vogel) occur in all of the biomes in Minas Gerais. Most of the species are listed by IUCN as being of least concern, except *A. angustifolia*, which is considered critically threatened and *P. lambertii*, which is near to threat/endangered. More information on the species can be seen in Table 4.2.

Table 4.1 List of threatened species in Minas Gerais, Brazil red-listed by IUCN and present in the lists of the Vegetation Monitoring System Project. Species ordered by red list status.

Code	Class	Order	Family	Genus	Species	Authority	Red List status*	Red List criteria**
152699	Magnoliopsida	Caryophyllales	Cactaceae	*Arrojadoa*	*rhodantha*	(Gürke) Britton & Rose	LC	
152594	Magnoliopsida	Caryophyllales	Cactaceae	*Cereus*	*hildmannianus*	K.Schum.	LC	
152911	Magnoliopsida	Caryophyllales	Cactaceae	*Cereus*	*jamacaru*	DC.	LC	
151744	Magnoliopsida	Caryophyllales	Cactaceae	*Leocereus*	*bahiensis*	Britton & Rose	LC	
46518	Magnoliopsida	Caryophyllales	Cactaceae	*Opuntia*	*monacantha*	Haw.	LC	
46508	Magnoliopsida	Caryophyllales	Cactaceae	*Pereskia*	*aculeata*	Mill.	LC	
46509	Magnoliopsida	Caryophyllales	Cactaceae	*Pereskia*	*grandifolia*	Haw.	LC	
62374	Magnoliopsida	Caryophyllales	Cactaceae	*Pilosocereus*	*aurisetus*	(Werderm.) Byles & G.D.Rowley	LC	
62375	Magnoliopsida	Caryophyllales	Cactaceae	*Pilosocereus*	*brasiliensis*	(Britton & Rose) Backeb.	LC	
40893	Magnoliopsida	Caryophyllales	Cactaceae	*Pilosocereus*	*floccosus*	Byles & G.D.Rowley	LC	
152415	Magnoliopsida	Caryophyllales	Cactaceae	*Pilosocereus*	*machrisii*	(E.Y.Dawson) Backeb.	LC	
62377	Magnoliopsida	Caryophyllales	Cactaceae	*Pilosocereus*	*pentaedrophorus*	(Cels) Byles & G.D.Rowley	LC	
46512	Magnoliopsida	Caryophyllales	Cactaceae	*Quiabentia*	*zehntneri*	(Britton & Rose) Britton & Rose	LC	
19892939	Magnoliopsida	Fabales	Fabaceae Faboideae	*Andira*	*fraxinifolia*	Benth.	LC	
19891561	Magnoliopsida	Fabales	Fabaceae Mimosoideae	*Acacia*	*piauhiensis*	Benth.	LC	
19892097	Magnoliopsida	Fabales	Fabaceae Mimosoideae	*Mimosa*	*bimucronata*	(DC.) Kuntze	LC	
19892593	Magnoliopsida	Fabales	Fabaceae Faboideae	*Platypodium*	*elegans*	Vogel	LC	
40896	Magnoliopsida	Caryophyllales	Cactaceae	*Pilosocereus*	*fulvilanatus*	(Buining & Brederoo) F.Ritter	NT	

"Table 4.1, conclusion."

Code	Class	Order	Family	Genus	Species	Authority	Red List status*	Red List criteria**
34086	Pinopsida	Pinales	Podocarpaceae	*Podocarpus*	*lambertii*	Klotzsch	NT	
151834	Magnoliopsida	Caryophyllales	Cactaceae	*Brasilicereus*	*phaeacanthus*	(Gürke) Backeb.	EN	A2ac
40858	Magnoliopsida	Caryophyllales	Cactaceae	*Pereskia*	*aureiflora*	F.Ritter	EN	A2c+4c
32975	Pinopsida	Pinales	Araucariaceae	*Araucaria*	*angustifolia*	(Bertol.) Kuntze	CR	A2cd
40888	Magnoliopsida	Caryophyllales	Cactaceae	*Pilosocereus*	*azulensis*	N.P.Taylor & Zappi	CR	B1ab(iii)
61926	Magnoliopsida	Fabales	Fabaceae Caesalpinioideae	*Dimorphandra*	*wilsonii*	Rizzini	CR	B1ab(ii,v)+ 2ab(ii,v); C2a(i,ii); D

*Red list status description: LC - least concern; NT - near threatened; EN - endangered; CR - critically endangered.
**Full description of red list criteria can be seen in http://jr.iucnredlist.org/documents/redlist_cats_crit_en.pdf; IUCN (2012).

Table 4.2 List of threatened species in Minas Gerais, Brazil selected for this study and their number of occurrences used to model the potential distribution. Species are sorted alphabetically by Genus.

Family	**Genus**	**Species**	**Authority**	**Red List status/criteria***	**Total records**
FABACEAE FABOIDEAE	*Andira*	*fraxinifolia*	Benth.	LC	136
ARAUCARIACEAE	*Araucaria*	*angustifolia*	(Bertol.) Kuntze	CR + A2cd	28
CACTACEAE	*Cereus*	*jamacaru*	DC.	LC	45
FABACEAE MIMOSOIDEAE	*Mimosa*	*bimucronata*	(DC.) Kuntze	LC	32
CACTACEAE	*Pereskia*	*aculeata*	Mill.	LC	23
CACTACEAE	*Pereskia*	*grandifolia*	Haw.	LC	46
FABACEAE FABOIDEAE	*Platypodium*	*elegans*	Vogel	LC	25
PODOCARPACEAE	*Podocarpus*	*lambertii*	Klotzsch	NT	142

*Red list status and criteria description: LC - least concern; NT - near threatened; EN - endangered; CR - critically endangered.Full description of red list criteria can be seen in http://jr.iucnredlist.org/documents/redlist_cats_crit_en.pdf; IUCN (2012).

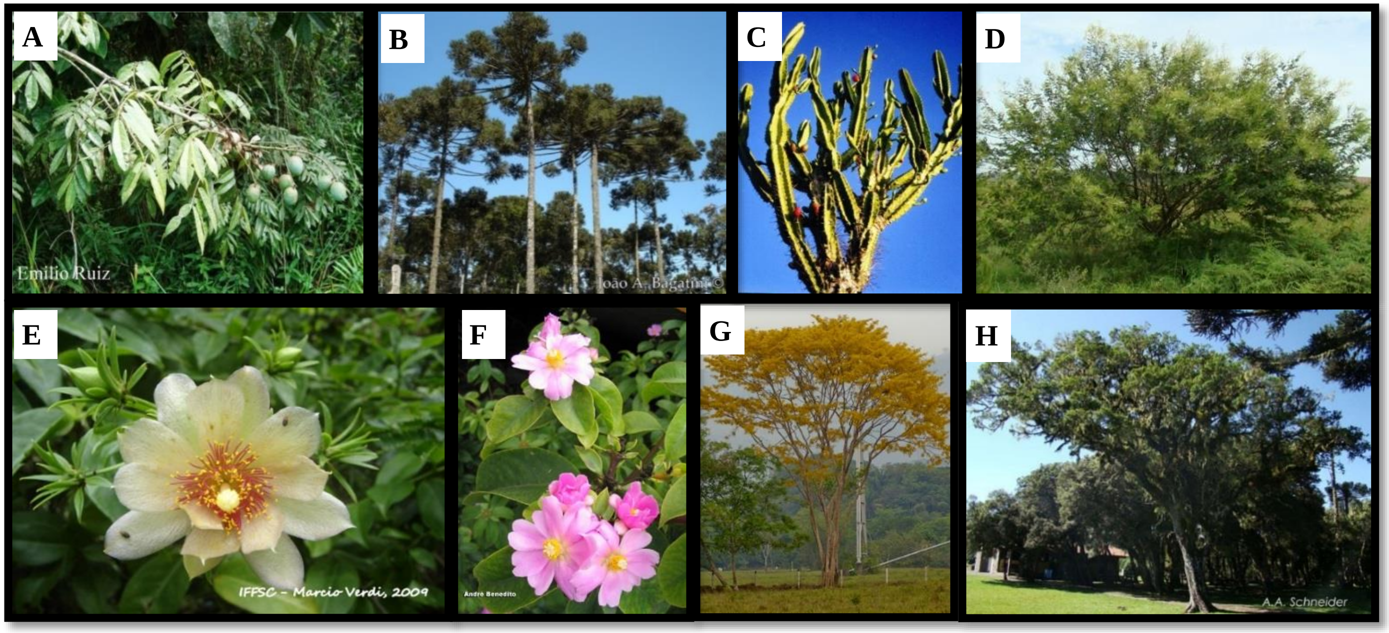
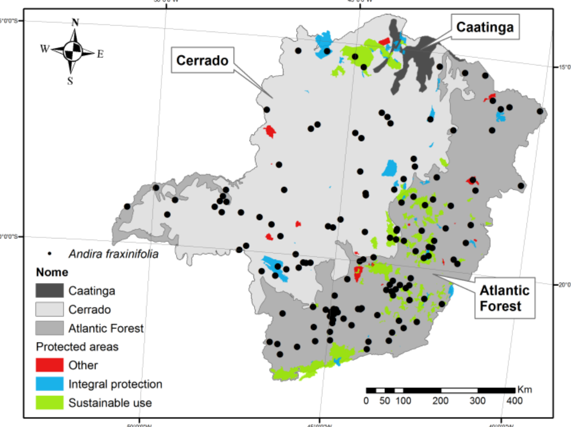
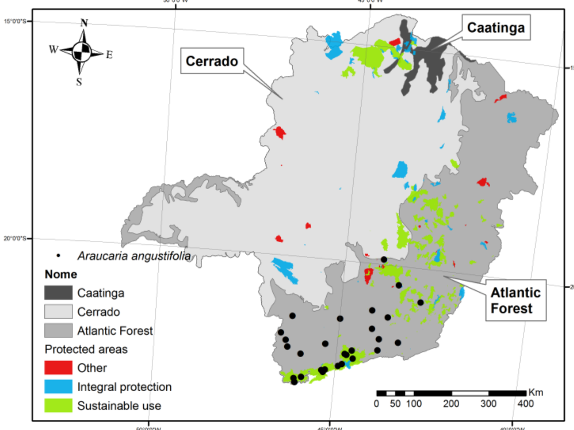


Figure 4.2 The eight threatened species selected for the study. A. *Andira fraxinifolia* (Benth.) (by Ruiz, E., available at http://sites.unicentro.br/wp/manejoflorestal/10113-2/); B. *Araucaria angustifolia* (Bertol.) Kuntze (by Bagatini, J.A., available at http://www.ufrgs.br/fitoecologia/florars/open_sp.php?img=14358); C. *Cereus jamacaru* DC (available at http://community.fortunecity.ws/greenfield/swallowtail/785/cereus_jamacaru.jpg); D. *Mimosa bimucronata* (DC.) Kuntze (by Scheineider, A.A., available at www.ufrgs.br); E. *Pereskia aculeata* Miller (By Verdi, M., available at www.ufrgs.br); F. *Pereskia grandifolia* Haw. (by Benedeto, A., available at http://www.jardimdesuculentas.net76.net/fichas/cac/pereskia_ grandifolia.html); G. *Platypodium elegans* Vogel (available at ibflorestas.org.br); H. *Podocarpus lambertii* Klotzsch (by Schneider,A.A., available at www.ufrgs.br).

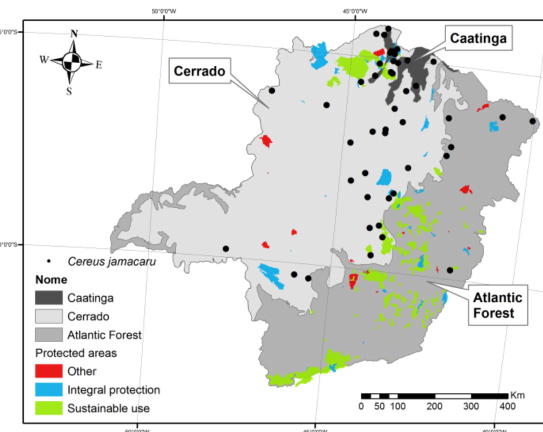
A


Caatinga
Cerrado
Atlantic Forest
Andira fraxinifolia
Nome
Caatinga
Cerrado
Atlantic Forest
Protected areas
Other
Integral protection
Sustainable use
0 50 100 200 300 400 Km

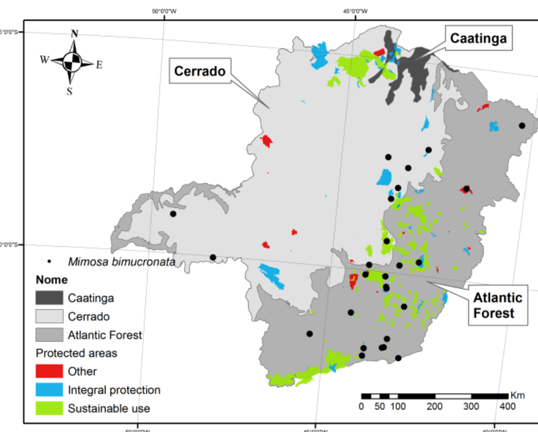
B


Caatinga
Cerrado
Atlantic Forest
Araucaria angustifolia
Nome
Caatinga
Cerrado
Atlantic Forest
Protected areas
Other
Integral protection
Sustainable use
0 50 100 200 300 400 Km

C


Caatinga
Cerrado
Atlantic Forest
Cereus jamacaru
Nome
Caatinga
Cerrado
Atlantic Forest
Protected areas
Other
Integral protection
Sustainable use
0 50 100 200 300 400 Km

D


Caatinga
Cerrado
Atlantic Forest
Mimosa bimucronata
Nome
Caatinga
Cerrado
Atlantic Forest
Protected areas
Other
Integral protection
Sustainable use
0 50 100 200 300 400 Km

E

F

G

H

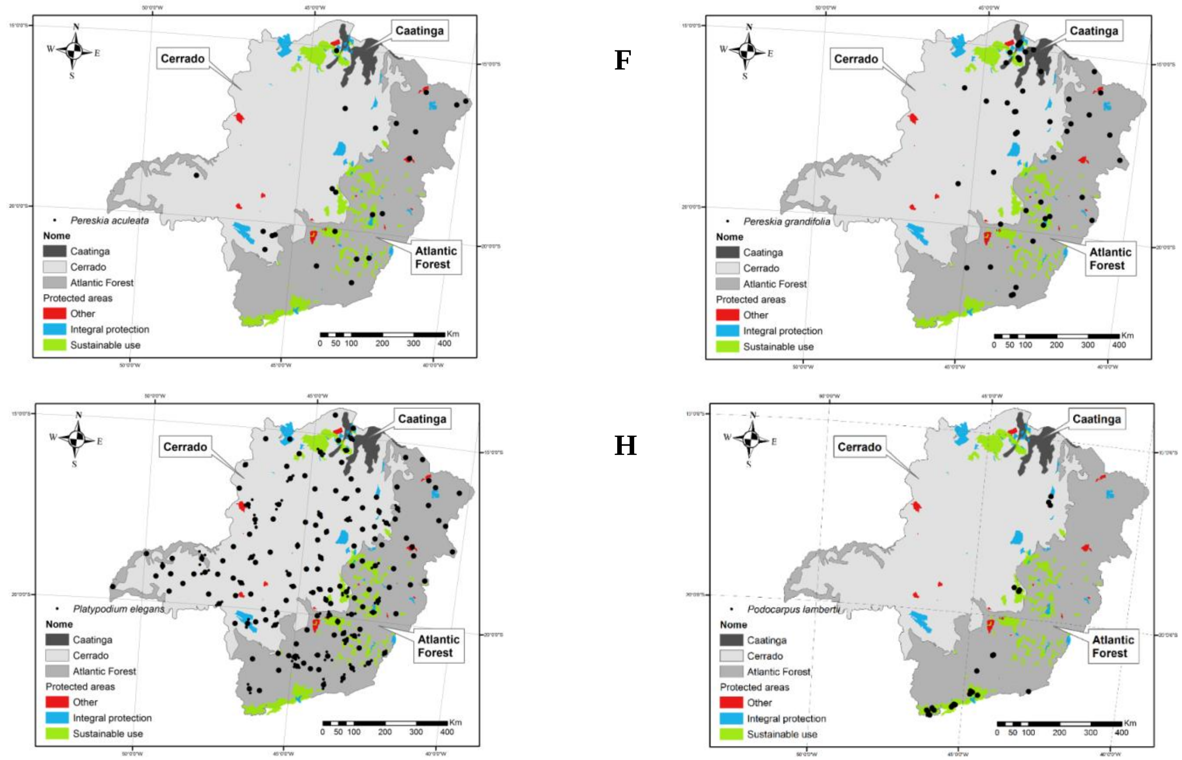


Figure 4.3 Occurrence points of the eight threatened species selected for the study. A. *Andira fraxinifolia* (Benth.); B. *Araucaria angustifolia* (Bertol.) Kuntze; C. *Cereus jamacaru* DC; D. *Mimosa bimucronata* (DC.) Kuntze; E. *Pereskia aculeata* Miller; F. *Pereskia grandifolia* Haw.; G. *Platypodium elegans* Vogel; H. *Podocarpus lambertii* Klotzsch. Three categories of protected areas (PAs) are also mapped (see descriptions in section *Environmental suitability and Priority areas for conservation*).

4.2.3 Environmental variables

We obtained 20 bioclimatic variables related to temperature, precipitation and topography (Table 4.3) from the WorldClim dataset (Hijmans et al. 2005; http://www.worldclim.org/), which consists of a set of global climate layers generated through interpolation of climate data from weather stations onto a 30″ grid (c. 1 km^2 resolution at the latitudes at which we were working). Eleven of the variables are temperature-derived, another eight are precipitation-derived and the last one is elevation (i.e. a Digital Elevation Model, DEM). We used the DEM data to generate slope values (in degrees) using the ‘Surface Analysis’ function in the GIS software package ArcGIS 10.1© (ESRI, Redlands, California, USA). We used another three variables related to soil characteristics (type, texture and organic matter) provided by the Ecological-economic Zoning of Minas Gerais Project – ZEE (Scolforo et al. 2008). All environmental variables were resampled to 1 km^2 resolution and a GCS projection and WGS84 datum. In total, we assembled 24 environmental parameters.

Table 4.3 Environmental variables selected to model the potential distribution of threatened species in Minas Gerais, Brazil, using the maximum entropy model. Data indicated as being sourced from "Worldclim" are from http://www.worldclim.org/bioclim.

Variable	Description	Source/Reference
BIO1	Annual Mean Temperature	Worldclim
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))	Worldclim
BIO3	Isothermality (BIO2/BIO7) (* 100)	Worldclim
BIO4	Temperature Seasonality (standard deviation *100)	Worldclim
BIO5	Max Temperature of Warmest Month	Worldclim
BIO6	Min Temperature of Coldest Month	Worldclim
BIO7	Temperature Annual Range (BIO5-BIO6)	Worldclim
BIO8	Mean Temperature of Wettest Quarter	Worldclim
BIO9	Mean Temperature of Driest Quarter	Worldclim

"Table 4.3, conclusion."

Variable	Description	Source/Reference
BIO10	Mean Temperature of Warmest Quarter	Worldclim
BIO11	Mean Temperature of Coldest Quarter	Worldclim
BIO12	Annual Precipitation	Worldclim
BIO13	Precipitation of Wettest Month	Worldclim
BIO14	Precipitation of Driest Month	Worldclim
BIO15	Precipitation Seasonality (Coefficient of Variation)	Worldclim
BIO16	Precipitation of Wettest Quarter	Worldclim
BIO17	Precipitation of Driest Quarter	Worldclim
BIO18	Precipitation of Warmest Quarter	Worldclim
BIO19	Precipitation of Coldest Quarter	Worldclim
Altitude	Elevation	Worldclim
Slope	Slope	by Authors
Soil_type	Soil type	ZEE-MG
Soil_text	Soil texture	ZEE-MG
Org_mat	Organic Matter	ZEE-MG

We removed redundancy from the set of predictor variables by testing for multicollinearity using cross-correlations (Pearson correlation coefficient, r) in ENMTools 1.3 software. This resulted in a matrix of pair-wise variable comparisons, where the variables were considered correlated when $r > 0.85$ (Zuur et al. 2007). We disregarded those variables that were strongly correlated and, therefore, deemed to be redundant. To do so, the selection criterion to choose one of them was an intuitive judgment on the potential biological relevance to the distribution of each species and the ease of interpretation. After removing correlated variables from the dataset, we ran a modelling test to analyse the performance of variables. We then removed variables that contributed less than 10 % to the model explanation for each species. Despite being part of the set of highly cross-correlated variables identified by our testing for multicollinearity, some variables were included in the model based on the same selection criterion used to remove the redundant variables (Table 4.4).

Table 4.4 Maximum entropy model results showing the percentage contribution of each environmental variable (abbreviations explained in Table 4.3) to explanation of the spatial distribution of each species.

	A. fraxinifolia		***A. angustifolia***		***C. jamacaru***		***M. bimucronata***		***P. aculeata***		***P. grandifolia***		***P. elegans***		***P. lambertii***	
AUC	0.67	**%**	0.962	**%**	0.756	**%**	0.778	**%**	0.646	**%**	0.772	**%**	0.526	**%**	0.939	**%**
Variable selected	bio4	27.2	bio9	53.2	bio14	53.1	bio3	73.7	soil_type	59.2	bio12	17.5	bio6	22.5	bio9	66.7
	bio3	18.4	bio11	20.7	soil_type	21.3	soil_type	13.7	bio3	27.8	bio15	15.8	bio4	14.8	bio1	16.3
	bio14	11.3	bio14	10.4	bio17	19.5	bio14	2.9	bio12	5.4	altitude	13.9	soil_type	13.1	slope	5.5
	soil_type	10.9	bio17	6.2	bio5	2.8	slope	2.6	bio4	4.5	bio3	12.5	slope	11.7	soil_type	3.6
	bio2	7.9	soil_type	4.9	bio15	2.2	bio2	2.4	bio17	1.8	bio14	11.9	bio3	10.7	bio5	2.5
	bio15	7.9	bio12	2	bio19	0.4	bio4	1.5	bio14	1.3	bio19	11.8	bio1	10.3	bio16	2.4
	slope	6.2	bio8	0.9	bio12	0.3	org_mat	1.5	bio16	0.1	bio5	9.1	bio14	8.7	bio8	1.8
	bio9	3.4	bio10	0.9	bio11	0.2	bio19	1.1			bio10	3.2	bio8	8.1	bio11	0.7
	bio17	2.6	bio1	0.8	bio18	0.1	bio11	0.4			bio6	2.2			bio6	0.4
	bio11	2	bio6	0.7			bio9	0.1			bio8	1.5			bio10	0.1
	bio6	1.2	bio5	0.3			bio1	0			bio18	0.5			bio12	0.1
	bio1	1									bio1	0.2				
											bio11	0.1				
											bio17	0.1				
											bio16	0				

4.2.4 Modelling procedure

To model the distributions of our species of interest, we used the well-known maximum entropy method, MaxEnt (Phillips et al. 2004, Phillips et al. 2006), which estimates the probability distribution for a species' occurrence based on environmental constraints (Phillips et al. 2006). MaxEnt is one of the most commonly used methods for inference of species distributions and environmental tolerances from occurrence data (Phillips et al. 2006), and has provided strong evidence of best performance when tested against many different methods (Elith et al. 2006), mainly because it is able to remain effective even when dealing with small sample sizes (Pearson et al. 2007, Elith et al. 2011). It requires only species presence and values for the environmental variables, which may be continuous or categorical. We used the freely-available MaxEnt version 3.3.3k (http://www.cs.princeton.edu/ ~schapire/maxent/) to generate estimates of the probability of species presence at each location within the study area for the eight threatened species we had previously selected.We followedthese steps: firstly, after obtaining the species occurrence points from the three different databases, we removed those points that felt within a 1 km buffer area to avoid spatial autocorrelation. We resampled and re-projected the environment variables, and we then tested them for multicollinearity, and removed redundancy. Next step was applying the modelling algorithm (MaxEnt) to the modelling dataset, partitioning it into training and test dataset. After performing the model calibration, and selecting the most suitable parameters/predictors for each species in our dataset, we used the training data to model and create the maps of the species current distribution. The final step was using the test data to validate models and assess the classification accuracy by testing the predictive performance through the AUC (area under the curve).

We used the default parameters in MaxEnt, with some slight changes. We used the "auto features" option, which allows the environmental layers to be used to produce "features" that constrain the probability distribution that is being computed according to the number of presence records for the species being modelled using general empirically-derived rules (Elith et al. 2011). MaxEnt currently has six feature classes: linear, product, quadratic, hinge, threshold and categorical (further details in Elith et al. 2011). We used the "logistic" output format because it provides output values, which are probabilities (between 0 or 1), scaled up in a non-linear way for easier interpretation (Phillips and Dudík 2008, Elith et al. 2011). We defined the maximum number of background points to 20,000 as the number of pixels in the study area was large. The background corresponds to a collection of points of the covariates from the landscape of interest (study area) with associated covariates, determining the distribution of covariates in the landscape (Elith et al. 2011).

We used the default value of 1 for the regularization multiplier. This parameter affects how focused or closely-fitted the output distribution is (Phillips and AT&T Research 2006). A value smaller than 1 will result in a more localized output distribution that is a closer fit to the given presence records, but can result in overfitting (fitting so close to the training data that the model does not generalize well to independent test data), while a larger regularization multiplier will give a more spread out, less localized prediction (Phillips and AT&T Research 2006). We set the maximum number of iterations to 5,000 and the convergence threshold to 0.0001 (Phillips and Dudík 2008). For each species, MaxEnt starts with a uniform distribution, and performs a number of iterations, each of which increases the probability of the sample locations for the species (Phillips and AT&T Research 2006). The probability is shown in terms of "gain", which starts at zero (the gain of the uniform distribution), and increases as MaxEnt increases the probabilities of the sample locations (Phillips and AT&T Research

2006). The gain increases iteration by iteration, until the change from one iteration to the next falls below the convergence threshold, or until maximum iterations have been performed (Phillips and AT&T Research 2006).

We used the cross-validation test as the replicated run type, which uses selected random points during the analysis. The number of replicates was set to 10. The "replicates" option can be used to do multiple runs for the same species (Phillips and AT&T Research 2006). The cross-validation test randomly splits the occurrence data into a number of equal-size groups called "folds", and models are created leaving out each fold in turn, so that the influence of the left-out folds can be evaluated (Phillips and AT&T Research 2006). Cross-validation has one big advantage over using a single training/test split: it uses all of the data for validation, thus making better use of small data sets (Phillips and AT&T Research 2006).

The threshold applied for environmental suitability was ‘10 percentile training presence’ (Liu et al. 2005), which discards those 10% of the records that have the lowest values of this index. We adopted this threshold as an extra precaution due to possible inaccuracies in geo-referenced data from different sources. To validate the models, we used the cross-validation method inspecting the area under the curve (AUC) value from the receiver operating characteristics curve (ROC), which measures the quality of a ranking of sites (Fielding and Bell 1997), and expert knowledge on the species distribution to evaluate the models. The AUC is the probability that a randomly selected presence site will be ranked above a randomly chosen absence site (Elith et al. 2011). A perfect ranking achieves the best possible AUC of 1.0, while a random ranking has on average an AUC of 0.5. Models with AUC higher than 0.75 are considered potentially useful (Elith 2002). When using presence-only models, the AUC is calculated using background data (called pseudo-absences, Elith et al. 2011). The interpretation of the AUC is then based on the probability that a randomly chosen presence site is ranked above a random background site (Phillips et al. 2006). Additionally, using presence-only data, we should interpret the grid cells with no occurrence localities as "negative examples" to use ROC curves, even if they support good environmental conditions for the species. In this case, the maximum AUC is therefore less than one, and is even smaller for wider-distributed species (Wiley et al. 2003). We used the jackknife estimator to assess the importance of each variable in the models and to determine how much exclusive information each variable provides alone or jointly with other variables (Elith et al. 2011, Phillips et al. 2006).

4.2.5 Environmental suitability and Priority areas for conservation

The outputs from the MaxEnt program were probabilistic maps of geographical distribution in ASCII format. These were converted into gridded data compatible with ArcGIS. We considered sites of low environmental suitability those with less than 50% probability of species presence, while sites with more than 50% probability of species presence were considered highly environmentally suitable locations (following Jimenez-Valverde et al. 2008). We then combined the potential distribution maps produced for each of the eight species into a single map in order to obtain the environmental suitability for all of the species together. We then calculated the environmental suitability at each location considering the presence of species that had occurrence probability values exceeding 50% in a given location. However, we found that a maximum of 6 species occurred together at any location. Therefore, the combined map consisted of values of environmental suitability that ranged from 0 – no occurrence of species, to 6 – occurrence of the six species at the same location. We overlaid the environmental suitability map with the protected area boundaries (CNUC/MMA 2015). This consisted of an ArcGIS shape file showing the locations of 246 protected areas in Minas Gerais, which fell into three categories (Figure 4.3). The first category contained 63 protected areas classified as being of "integral protection" (where only restricted use is allowed, e.g. research). The second contained 159 protected areas, which are designated as being of "sustainable use" (which allows some types of intervention, e.g. collection and use of natural resources; privately-owned protected areas fall into this category). The final category contained 24 protected areas, which are classified as being under other types of protection (e.g. indigenous areas), about which we found no usage information (National Protected Areas System – Brazil 2000). The product of this procedure was a map containing the environmentally suitable areas for the eight species combined within the PAs. The next step consisted of overlaying the map of environmentally suitable areas for the eight species combined, within the PAs, with the existing natural vegetation remnants of Cerrado, Caatinga and Atlantic Forest. These remnants were obtained from a land use map of 2011 provided by the Vegetation Monitoring System of Minas Gerais (Carvalho and Scolforo – unpublished data). The product obtained from this overlay procedure was a map containing the environmentally suitable areas for the eight species combined within the PAs that had not been converted into other land use/cover, but remain as natural vegetation fragments. In summary, we quantified environmentally suitable areas for threatened species within the PAs, and we also identified environmentally suitable areas after overlaying them with the existing natural vegetation remnants, giving a more realistic idea about their occurrence and protection status. This also allowed us to identify the gaps in protection, which consisted of those sites where the species environmental suitability was high but there is currently no PA. These sites were considered as priority areas for conservation.

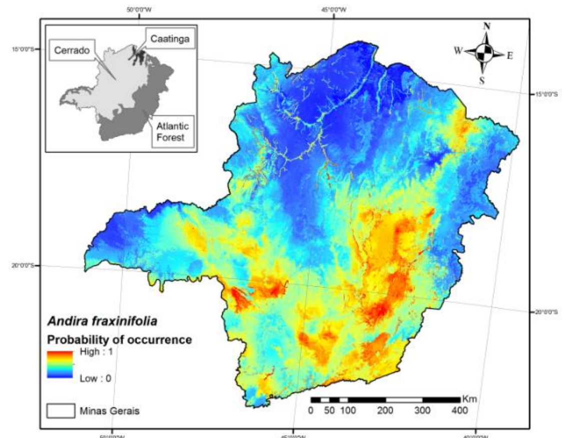
4.3 Results

4.3.1 Model accuracy and species potential distribution

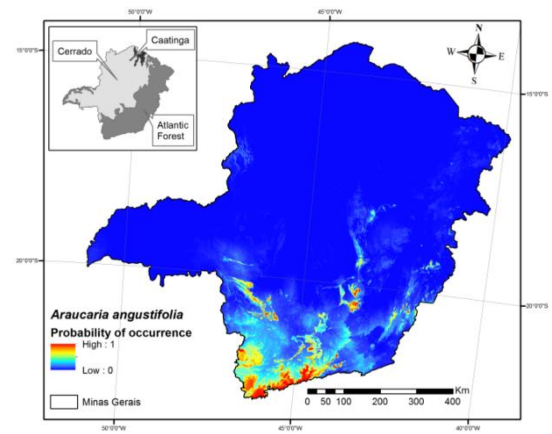
The accuracy of the models varied between the species; their AUC values ranged from 0.526 to 0.962. We observed that species with higher number of occurrence points (namely *A. fraxinifolia* and *P. elegans*) had the lowest values of AUC. A variety of environmental variables appeared in the habitat suitability models with varying importance for the species studied (Table 4.4). Considering only the most important variable in each model, the only consistency observed was that in the models for both of the species belonging to the order Pinales, *A. angustifolia* and *P. lambertii*, the mean temperature of the driest quarter of the year was the most important variable. The models for these two species were also similar in terms of their list of variables, sharing nine of the eleven variables in each model. The models for species from the other two orders were less similar. The three species from the order Fabales shared six environmental variables in total, while the number of variables used to build their models ranged between 8 and 12. The most important variables in their models were those that related to temperature, namely the temperature seasonality (for *A. fraxinifolia*), the isothermality (for *M. bimucronata*), and the minimum temperature of coldest month (for *P. elegans*). Species from the order Caryophyllales presented the lowest similarity. The three species in this group only shared three variables among their models. The most important variables in the models of *C. jamacaru*, *P. aculeate*,*P. grandifolia* were precipitation in the driest month, soil type, and annual precipitation respectively. The contribution of the single most important environment variable to each model ranged from 17.5 to 73.7% of the variance explained (Table 4.4).

The distribution of suitable sites also varied between species, showing that some species are more environmentally restricted, while others have a wider distribution across the biomes in Minas Gerais (Figure 4.4). According to the models, species from the order Pinales (*A. angustifolia,* Araucariaceae; and *P. lambertii,* Podocarpaceae) are restricted to montane and sub-montane forests located in the higher altitudes of the Atlantic Forest in a region called *Serra da Mantiqueira* in Southern Minas Gerais. The potential distribution of four other species was higher in the core of the Atlantic Forest of Minas Gerais, two of them belonging to the family Fabaceae Faboideae, another to the family Fabaceae Mimosoideae, and the last one to the family Cactaceae (*A. fraxinifolia*, *P. elegans, M. bimucronata,* and *P. aculeata,* respectively). *P. aculeata* appears to be more restricted to the dry deciduous forests of Caatinga and Atlantic Forest in the North of Minas Gerais.

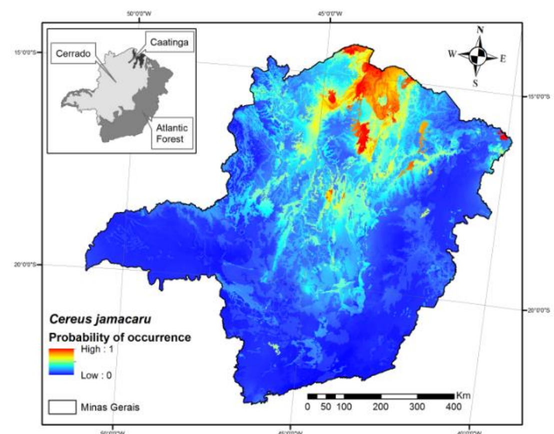
A


Cerrado
Caatinga
Atlantic Forest
Andira fraxinifolia
Probability of occurrence
High : 1
Low : 0
Minas Gerais
0 50 100 200 300 400 Km

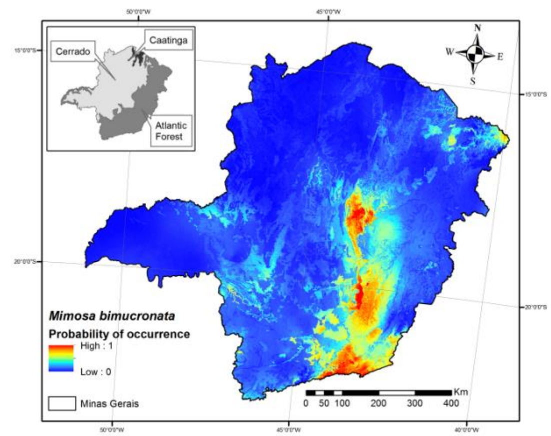
B


Cerrado
Caatinga
Atlantic Forest
Araucaria angustifolia
Probability of occurrence
High : 1
Low : 0
Minas Gerais
0 50 100 200 300 400 Km

C


Cerrado
Caatinga
Atlantic Forest
Cereus jamacaru
Probability of occurrence
High : 1
Low : 0
Minas Gerais
0 50 100 200 300 400 Km

D


Cerrado
Caatinga
Atlantic Forest
Mimosa bimucronata
Probability of occurrence
High : 1
Low : 0
Minas Gerais
0 50 100 200 300 400 Km

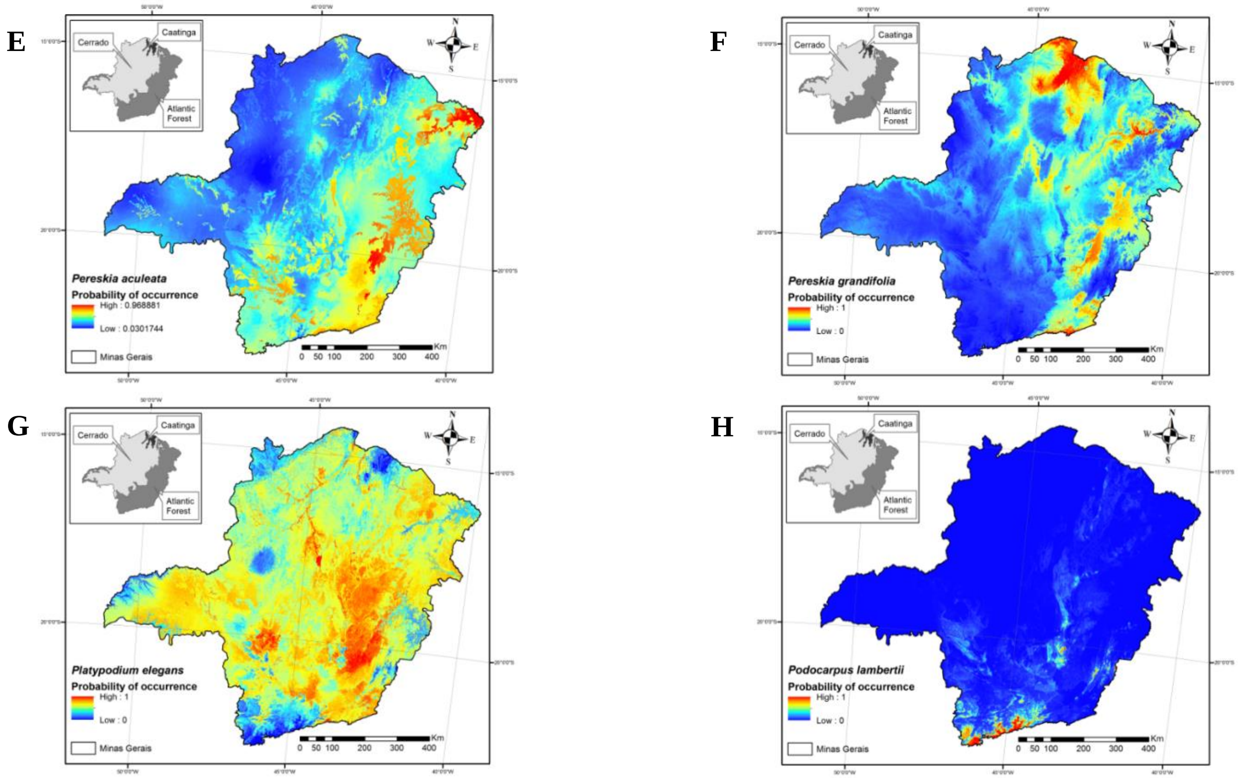
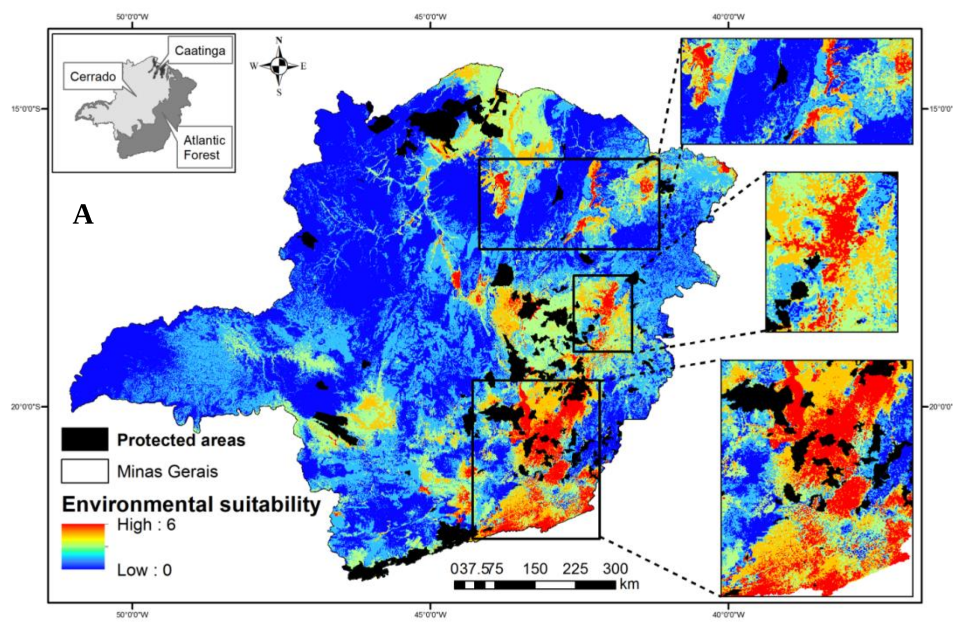
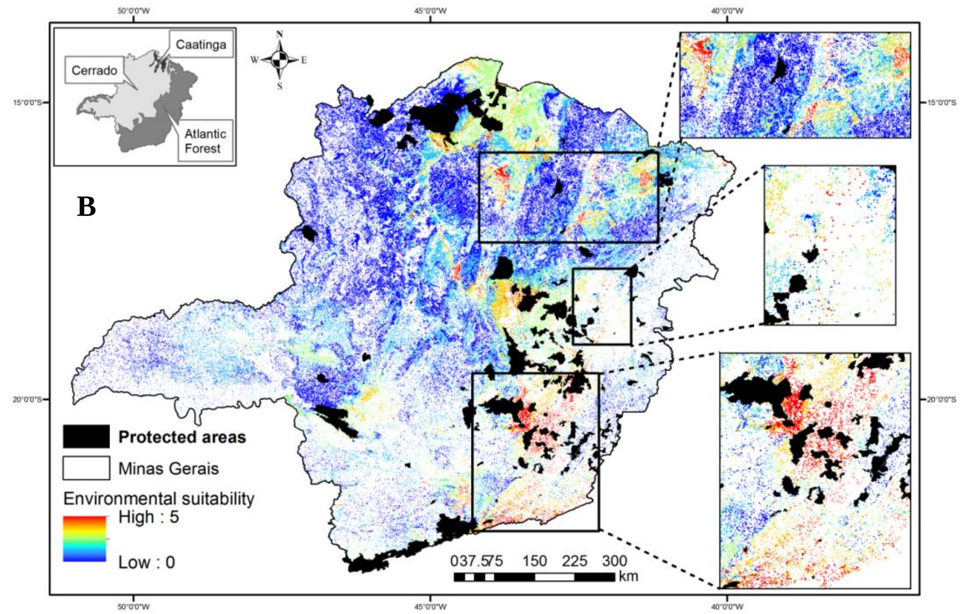


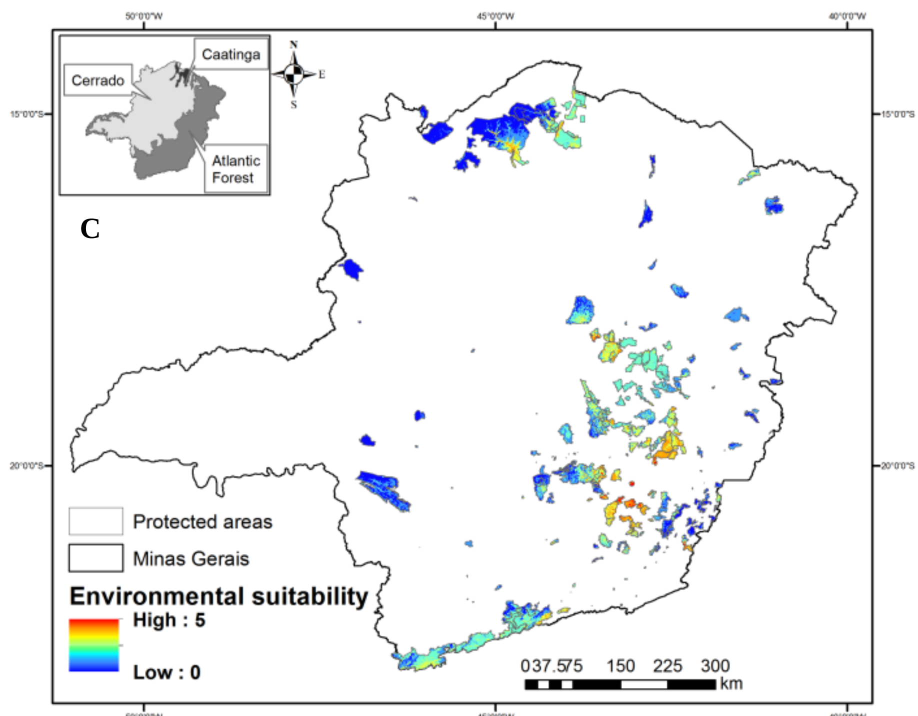
Figure 4.4 Probability of occurrenceof the eight threatened species selected for the study. A. *Andira fraxinifolia* (Benth.); B. *Araucaria*angustifolia (Bertol.) Kuntze; C. Cereus jamacaru DC; D. Mimosa bimucronata (DC.) Kuntze; E. Pereskia aculeata Miller; F. Pereskia grandifolia Haw.; G. Platypodium elegans Vogel; H. Podocarpus lambertii Klotzsch.

4.3.2 Environmental suitability and priority areas for conservation

The map of environmental suitability for all species combined shows that a large majority of the suitable sites are located in the Atlantic Forest biome, a smaller number are in the Caatinga and fewest are in the Cerrado (Figure 4.5A). This map also shows that approximately 32.6 million hectares (55%) of Minas Gerais State are favourable for the occurrence of at least one of the species studied (Table 4.5). Table 4.5 shows the figures of the extent of environmentally suitability areas ranging from no species occurrence up to locations with the presence of six species together, considering the area of Minas Gerais (MG) and of PAs, the area of natural vegetation remnants within MG and within PAs. This table also shows some area-ratio relations, e.g. the relation between MG-remnants and MG-area showing the area-ratio between the areas of natural vegetation remnants in comparison to the state area. Similar relations are PA-remnants and PA-area – area-ratio between the areas of natural vegetation remnants within PAs in comparison to the PAs area, and the PA-remnants and MG-area – area-ratio between the areas of natural vegetation remnants within PAs in comparison to the state area.Furthermore, the highest environmental suitability values generated by the model represent the potential occurrence of a maximum of six species at any one location, and a maximum of five species at locations that are still covered by natural vegetation remnants within the PA’s boundaries (Figure 4.6). Figure 4.6 A e B show the suitable or unsuitable areas for the threatened species considering a ranking of occurrence of these species combined, ranging from 0 (unsuitable) to 6 (suitable for 6 species). Figure 4.6 A shows the percentages for the environmentally suitable areas for the whole State and PAs, and the relation between PA/MG shows area-ratio between the area of PAs in comparison to the State’s area. Figure 4.6 B shows the percentages for the environmentally suitable areas that are covered by natural vegetation in the State and within PAs. The relation PA-remnants/MG-remnants show the area-ratio between the areas of natural vegetation remnants within PAs in comparison to areas of natural vegetation remnants in MG. In addition, a majority of sites identified by the model as suitable for the species studied are not available as natural habitats anymore, because they have been converted to other land uses, according to the land use map from the Vegetation Monitoring System of Minas Gerais (Carvalho and Scolforo – unpublished data). Thus, the actual area of suitable sites that were still natural habitat totalled less than a third (approximately 9.6 million hectares) of the area indicated as suitable for at least one species (32.6 million hectares).




Caatinga
Cerrado
Atlantic Forest
B
Protected areas
Minas Gerais
Environmental suitability
High : 5
Low : 0
0 37.5 75 150 225 300
km


Caatinga
Cerrado
Atlantic Forest
C
Protected areas
Minas Gerais
Environmental suitability
High : 5
Low : 0
0 37.5 75 150 225 300
km

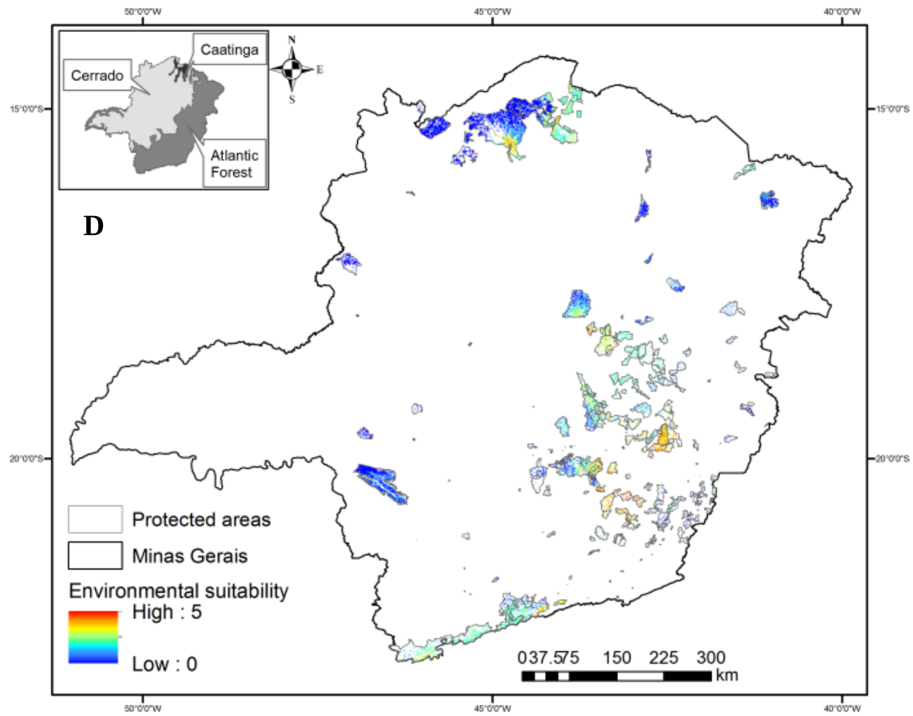


Figure 4.5 The environmental suitability for all threatened species combined for: A) Minas Gerais State (MG); B) Remnants of natural vegetation in MG; C) Protected areas (PA) within MG; and D) Remnants of natural vegetation in the PAs of MG. Inset maps show the sites of high potential occurrence of the most threatened species combined that are currently unprotected, and where new protected areas can be created (for those locations areas of existing natural vegetation remnants) or for setting of priorities to restore naturalhabitat such as replanting projects (where forest was converted to other land uses).

Table 4.5 The extent of environmentally suitability areas in hectares and as a percentage of the total area of Minas Gerais State and the protected areas within its boundaries. The environmental suitability ranking indicates the number of species for which the maximum entropy model found an area to be environmentally suitable. MG = Minas Gerais; PA= Protected areas; MG-remnants = natural vegetation remnants in MG; PA-remnants = natural vegetation remnants in the PAs within MG. The relation MG-remnants/MG-area shows area-ratio between the area of natural vegetation remnants in comparison to the state area, while the relation PA-remnants/PA-area show the area-ratio between the areas of natural vegetation remnants within PAs in comparison to the PAs area. The relation PA-remnants/MG-area shows the area-ratio between the areas of natural vegetation remnants within PAs in comparison to the state area.

Environmental suitability ranking	**MG**		**PA**		**MG-remnants**		**MG-remnants/ MG-area**	**PA-remnants**		**PA -remnants/ PA-area**	**PA-remnants/ MG-area**
	Area (ha)	**%**	**Area (ha)**	**%**	**Area (ha)**	**%**	**%**	**Area (ha)**	**%**	**%**	**%**
0	26,040,043.20	44.40	1,114,086.77	29.19	8,748,134.00	47.80	14.92	611,763.60	33.19	16.0312	1.04
1	17,526,678.02	29.88	825,141.38	21.62	4,978,818.00	27.20	8.49	359,013.50	19.48	9.40791	0.61
2	8,623,382.49	14.70	1,066,647.21	27.95	2,822,845.00	15.42	4.81	505,503.10	27.42	13.2467	0.86
3	4,202,180.46	7.17	525,264.96	13.76	1,214,030.00	6.63	2.07	239,515.00	12.99	6.27646	0.41
4	1,886,093.15	3.22	257,856.18	6.76	478,202.40	2.61	0.82	119,064.80	6.46	3.12008	0.20
5	368,967.70	0.63	27,085.93	0.71	61,276.89	0.33	0.10	8,486.83	0.46	0.2224	0.01
6	158.95	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	0.00
Total	58,647,503.97	100.00	3,816,082.42	100.00	18,303,306.41	100.00	31.21	1,843,346.83	100.00	48.3047	3.14

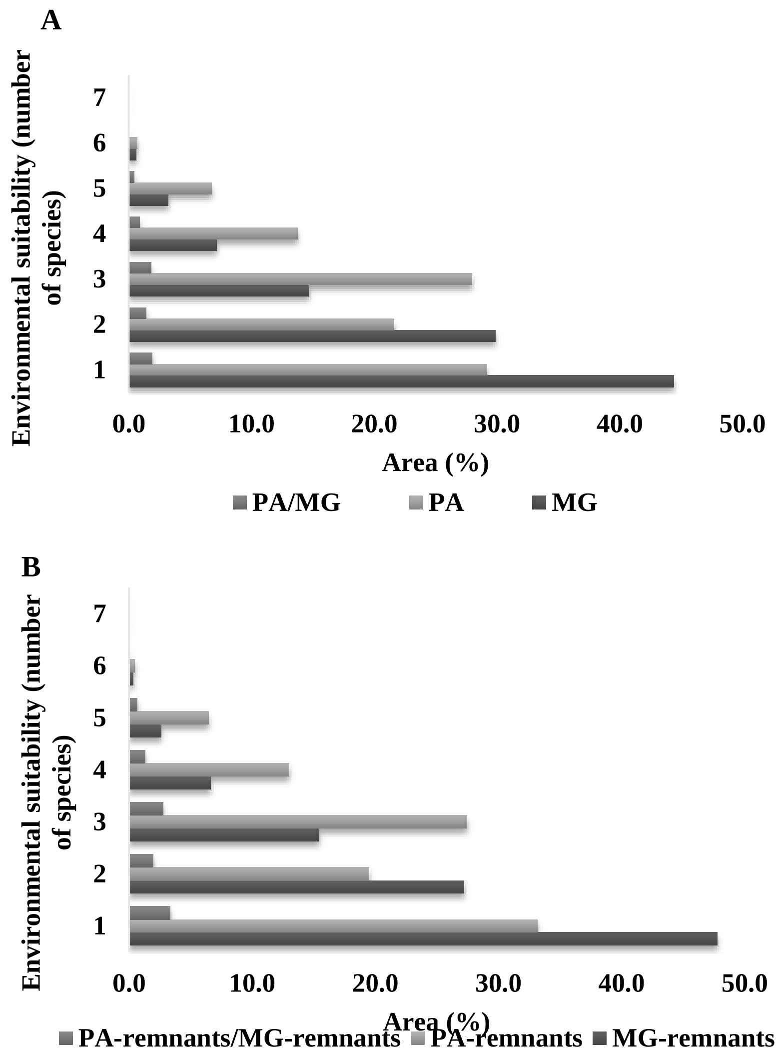


Figure 4.6 The area of environmental suitability in percentage based on the occurrence of threatened species for Minas Gerais State and for the protected areas within its boundaries. A) Environmental suitability considering the whole area of Minas Gerais; B) Environmental suitability considering remnants only. MG = Minas Gerais; PA= Protected areas. The relation PA/MG shows area-ratio between the area of PAs in comparison to the state area, while the relation PA-remnants/MG-remnants show the area-ratio between the areas of natural vegetation remnants within PAs in comparison to areas of natural vegetation remnants in MG.

We verified that PAs from all three categories (integral protection, sustainable use, and other types of protection) protect an area of 4.1 million hectares, which corresponds to approximately 7% of Minas Gerais. However, when we count only areas of existing natural vegetation remnants within the PAs, this area drops to 3.14% of Minas Gerais (1.8 million hectares), less than half of the PAs' combined area.

The map of environmental suitability for all species combined also showed significant gaps in the protection offered by the PA network, regarding the protection of the species studied. Considering that the area under protection in Minas Gerais represents less than 7% of its territory and the area of natural vegetation remnants within the PAs correspond only to 3.14%, if we focus on sites where at least one threatened species occurs, the total suitable area under protection is reduced to an even smaller percentage (Figure 4.5 C and D). However, we identified at least three sites of high potential occurrence of the most threatened species that are currently unprotected: one site located in Southern Minas Gerais; a site in the central Atlantic Forest; and the third site located in the North of Minas Gerais (Figure 4.5A and B).

4.4 Discussion

4.4.1 Model accuracy and species potential distribution

Considerable variation was present in the maximum entropy models. As expected, wider-ranging species (i.e. those with more occurrence points) had the lowest AUC values. A reason for this is precisely their wider distribution (Wiley et al. 2003). The widely-distributed species occupy a diverse set of environmental conditions which reduces the power of the MaxEnt modelling to predict species distributions based on environmental variation. The species from the order

Fabales, except for *M. bimucronata*, presented the most widespread distributions in Minas Gerais, and consequently the lowest values of AUC. Nevertheless, we kept these models in our analysis, since they showed consistent species distributions, as also observed by Anderson et al. (2003), and provided important information for their conservation. Species from the order Pinales showed the most restricted distributions, which were closely related to higher altitudes. This is consistent with the literature, since both *Araucaria angustifolia* and *Podocarpus lambertii* are indicator species of high altitudes (Veloso et al. 1991, Oliveira-Filho and Fontes 2000). They also had the highest values of AUC. Species from the family Cactaceae had an intermediate distribution, and consequently intermediate values of AUC.

We expected that a variety of environmental variables would be important to explain the potential distribution of species, as we worked with a range of species from different orders. We also expected some similarity among species within the same order or family, but that was not always the case. We did observe a strong similarity in the distributions of species from the order Pinales, which reflects the characteristics they have in common, both of them being indicators of high altitudes, for example (Oliveira-Filho and Fontes 2000). In addition, the most important variable determining their distribution was the mean temperature of the driest quarter, which implies that these species might be particularly sensitive to relatively high temperatures in the dry season. They are associated with higher altitude sites, where the mean annual temperature is lower than 20º C (Lopes et al. 2013). These two species also occur together in other regions of the Atlantic Forest in Southeast and Southern Brazil, as well as in Argentina (Farjon 2013, Thomas 2013).

Species from the order Fabales appear to have less in common than the species from the order Pinales, as they shared a smaller number of environmental variables among their models. However, while the main environmental variables determining their distribution are different for each species model, all of them are temperature-derived variables. *A. fraxinifolia* is more sensitive to temperature seasonality, which is the standard deviation of the temperature multiplied by 100 (Worldclim; Hijmans et al. 2005), and suggests this species cannot tolerate wide variations in temperature or extreme temperatures. Similarly, isothermality seems to regulate strongly *M. bimucronata* distribution, which means that this species requires the temperature to vary relatively little within individual months relative to its variation across the whole year (Worldclim; Hijmans et al. 2005, Carvalho 2004). Lastly, *P. elegans* appears to be regulated by the minimum temperature of the coldest month: despite having widespread distribution, this species seems not to be able to tolerate very low temperatures.

Species from the family Cactaceae showed the lowest similarity with regard the environmental variables shared in their models among all species studied. Although two of them belong to the same genus, *P. aculeata* and *P. grandifolia*, their distributions were influenced mainly by two very different variables: soil type and annual precipitation.*C. jamacaru* was also influenced most strongly by a precipitation-derived variable: the precipitation of the driest month. *C. jamacaru* is typical of dry deciduous forests of the Caatinga, in the Agreste (the ecotone between Caatinga and Atlantic Forest), and the Restinga, a mosaic of different coastal vegetation types, ranging from open scrub to forest (Henriques et al. 1986), which occurs in a narrow band between the sea and the Atlantic forest (Oliveira-Filho and Fontes 2000), and the Cerrado (Taylor and Zappi 2004). Although this species is tolerant to dry seasons, precipitation seems to be a limiting factor to its distribution.

4.4.2 Environmental suitability and priority areas for conservation

We have shown that the amount of land under protection in Minas Gerais is far from being sufficient for adequate conservation of the species studied, through the comparison of the environmental suitability map with maps of natural vegetation remnants and PA boundaries. The main gaps in protection of areas of the highest environmental suitability for these species are located in the Atlantic Forest biome, which is also the biome with the highest number of PAs and the largest area under protection. In addition, all of the sites identified as being of high suitability are within the Atlantic Forest boundaries in Minas Gerais. This result may be related to the geographic distribution of the species considered in this study, as they all occur in the Atlantic Forest, while only few of them also occur in the Caatinga and Cerrado, but none of them is exclusive to the latter two biomes. Additionally, the Atlantic Forest has also been focussed upon in several studies, so it is possible that many species from the Cerrado and Caatinga are not in the IUCN red lists due to the lack of knowledge about them. As the Cerrado and Caatinga have only a few PAs in Minas Gerais, they should also be prioritised for conducting inventories and creating natural reserves.

The information produced by this study is timely and highly relevant given the potential threats to the habitats of the species studied here and to overall biodiversity in Minas Gerais biomes due to anthropogenic actions (Groom 2012, Lopez-Poveda 2012a,b, Braun et al. 2013, Farjon 2013, Taylor et al. 2013a, Taylor et al. 2013b, Thomas 2013). Considering the huge losses incurred from the original land cover at the time of European arrival in South America in the 16th century, the current state of degradation of the three biomes studied here and the lack of knowledge in some regions (mainly in the Cerrado and Caatinga), the maximum entropy modelling approach is shown by the present study to be a useful tool for indicating environmentally suitable areas for reintroduction of species in biodiversity restoration projects. Environmental suitability maps for threatened species, such as those presented here, can help land use planning and management around their existing populations, discovery of new populations, identification of top-priority survey sites, or setting of priorities to restore natural habitat such as aforestation projects. The modelling approach used here could also be applied to other threatened species, including fauna as well as flora. In particular, it has the potential to aid research that is needed to gain a better understanding of threatened species in the Cerrado and Caatinga.

4.5 Conclusions

This work demonstrates the success of the application of maximum entropy modelling to determination of potential distributions of threatened plant species in the Atlantic forest, Cerrado and Caatinga biomes within the state of Minas Gerais, Brazil. The outcomes of this modelling provide important information for the implementation of conservation efforts within these biomes. Our results support the development of conservation for the species studied, since they are IUCN red-listed species. Furthermore, we identified significant gaps in protection for these species, implying that the number and effectiveness of protected areas suitable for the species is currently less than adequate for their efficient conservation. These gaps occur mainly in the Atlantic Forest, in three specific locations in Southeast, Central and Northern Minas Gerais. This suggests that these areas should be considered as priorities for conducting inventories of species distributions and proposal of new natural reserves. Further research should also be prioritised in the Cerrado and Caatinga biomes, since relatively little is known about the current and potential distributions of threatened species in both.

Acknowledgements

We would like to thank the Federal University of Lavras (UFLA) and for providing the Inventário Florestal de Minas Gerais databases, the SpeciesLink network, and Dr. Oliveira-Filho, AT (NeoTropTree databases) for providing the data. We also would like to thank Dr. Kamino, LHY, Dr. Loyola, RD, Dr. Bernardi, L and Dr. Lemes, P. for their valuable comments and advice on modeling procedures. L. Zanella acknowledges the support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), which provided a PhD scholarship.

REFERENCES

Anderson RP, Lewc D, Peterson T. 2003 Evaluating predictive models of species’distributions: criteria for selecting optimal models Ecological Modelling, 162:211-232.

Araújo MB, Williams PH, Fuller RJ. 2002. Dynamics of extinction and the selection of nature reserves Proceedings of the Royal Society B, Biological Sciences, 269:1971-1980.

Baillie JEM, Hilton-Taylor C, Stuart SN. 2004. IUCN Red List of Threatened Species A Global Species Assessment, Gland, Switzerland and Cambridge, UK, IUCN.

Biodiversistas. 2005. Revisão da Lista da Flora Brasileira Ameaçada de Extinção Available online: http://wwwbiodiversitasorgbr/floraBr/defaultasp (Accessed November 2012) Fundação Biodiversistas.

Bourg NA, McShea WJ, Gill DE. 2005. Putting a CART before the search: successful habitat prediction for a rare forest herb Ecology, 86:2793–2804.

Braun P, Machado M, Taylor NP. 2013. *Cereus jamacaru*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. (Downloaded on May 2015).

Brazil. 2000. Law N. 9985/00 National System of Nature Conservation Unit – SNUC Brasília Available from: http://wwwmmagovbr/port/sbt/dap/doc/snucpdf (Accessed March 2015).

Brooks TM, Mittermeier RA, da Fonseca GAB, Gerlach J, Hoffmann M, Lamoreux JF, Mittermeier CG, Pilgrim JD, Rodrígues ASL (2006) Global biodiversity conservation priorities. Science 313:58–61.

Butchart SHM et al. 2010 Global Biodiversity: indicators of recent declines Science, 328:1164-1168.

Carvalho LMT, Scolforo JRS. Unpublished data. Inventário Florestal de Minas Gerais: Monitoramento da Flora Nativa 2007-2009; 2009-2011.

Carvalho, PER. 2004. Circular Técnica 94: Marica - *Mimosa bimucronata* Colombo: EMBRAPA.

Castelletti CHM, Silva JMC, Tabarelli M, Santos AMM. 2004. Quanto ainda resta da Caatinga? Uma estimativa preliminar. In: Silva JMC, Tabarelli M, Lins L (eds). Biodiversidade da Caatinga: áreas e ações prioritárias para a conservação. Ministério do Meio Ambiente, Brasília, pp. 92-100.

CNUC/MMA. 2015. Cadastro Nacional de Unidades de Conservação. Ministério do Meio Ambiente. Available from www.mma.gov.br. (Accessed February 2015).

Dirzo R, Young HS, Galetti M, Ceballos G, Isaac NJB, Collen B. 2014. Defaunation in the Anthropocene Science 345:401-406.

Elith J, et al. 2006. Novel methods improve prediction of species' distributions from occurrence data Ecography 29:129-151.

Elith J, Phillips SJ, Hastie T, Dudík M, Chee Y, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists Diversity and Distributions, 17:43-57.

Elith J. 2002. Quantitative methods for modeling species habitat: Comparativeperformance and an application to Australian plants. In: Ferson S and Burgman M (Eds.). Quantitative methods for conservation biology. New York: Springer-Verlag.

Engler R, Guisan A, Rechsteiner L. 2004. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data J Appl Ecol 41: 263-274.

Fahrig L, Rytwinski T. 2009. Effects of roads on animal abundance: an empiricalreview and synthesis. Ecology and Society 14: 21.

Farjon A. 2013. *Podocarpus lambertii*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. (Downloaded on May 2015).

Ferrier S, Watson G, Pearce J, Drielsma M. 2002. Extended statistical approaches to modeling spatial pattern in biodiversity: the north-east New South Wales experience I Species-level modeling Biodiv Conserv 11: 2275–2307.

Ferro VG, Lemes P, Melo AS, Loyola R. 2014. The Reduced Effectiveness of Protected Areas under Climate Change Threatens Atlantic Forest Tiger Moths. PLoS One 9:e107792.

Fielding AH, Bell JF. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24: 38-49.

Gaston KJ. 1996. Species richness: measure and measurement Biodiversity: a biology of numbers and difference. Oxford: Blackwell Science, pp 77–113.

Groom A. 2012. *Platypodium elegans*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. (Downloaded on May 2015).

Guisan A, Thuiller W. 2005. Predicting species distribution: offering more than simple habitat models Ecol Lett, 8:993-1009.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology Ecol Modell 135:147-186.

Hijmans RJ, Cameron SE, Parra JL, Jones P, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas International Journal of Climatology, 25:1965-1978.

IBGE. 2004. Mapa de Biomas e de Vegetação. Available from ftp://ftpibgegovbr/Cartas_e_Mapas/Mapas_Murais/ (Accessed August 2014).

IUCN. 2012. IUCN Red List Categories and Criteria: Version 3.1. Second edition. Gland,Switzerland and Cambridge, UK: IUCN. Available from www.iucnredlist.org/technical-documents/categories-and-criteria. (Accessed Mar 2015).

IUCN. 2014. The IUCN Red List of Threatened Species Version 20143 Available from: http://wwwiucnredlistorg (Accessed November 2014).

Jiménez-Valverde A, Lobo J, Hortal J. 2008. Not as good as they seem: the importance of concepts in speciesdistribution modelling. Diversity and Distributions 14:885–890.

Lemes P, Melo AS, Loyola RD .2014. Climate change threatens protected areas of the Atlantic Forest Biodivers Conserv 23:357–368.

Liu C, Berry PM, Dawson TP, Pearson RG. 2005. Selecting thresholds of occurrence in the prediction of species distributions Ecography, 28:385-393.

Lopes PMO, Valeriano DM, Silva BB, Moura GBA, Silva AO. 2013. Simulação do saldo de radiação na Serra da Mantiqueira. Revista Brasileira de Engenharia Agrícola e Ambiental. 17(7):780–789.

Lopez-Poveda L. 2012a. *Andira fraxinifolia*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. (Downloaded on May 2015).

Lopez-Poveda L. 2012b. *Mimosa bimucronata*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. (Downloaded on May 2015).

Margules CR, Pressey RL. 2000. Systematic conservation planning. Nature 405:243-253.

Mittermeier R, Turner W, Larsen F, Brooks T, Gascon C. 2011. Global Biodiversity Conservation: The Critical Role of Hotspots. Berlin: Springer Berlin Heidelberg.

Mittermeier RA, Gil RP, Hoffman M, Pilgrim J, Brooks T, Mittermeier CG, Lamoreux J, Fonseca GAB. 2005. Hotspots revisited: earth's biologically richest and most endangered terrestrial ecoregions. 2nd ed. Boston: University of Chicago Press.

MMA – Ministério do Meio Ambiente. 2008. Instrução Normativa no 6, de 23 de setembro de 2008.

Oliveira-Filho AT, Fontes MAL. 2000. Patterns of floristic differentiation among Atlantic Forests in Southeastern Brazil and the influence of climate. Biotropica 32:793-810.

Pearson RG, et al. 2007. Predicting species' distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar Journal of Biogeography 34:102-117.

Phillips S, AT&T Research. 2006. A brief tutorial on MaxEnt. Available from: https://www.cs.princeton.edu/~schapire/maxent/ (Accessed on January 2015).

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions Ecological Modelling 190:231-259.

Phillips SJ, Dudik M . 2008. Modeling of species distributions withMaxent: new extensions and a comprehensive evaluation. Ecography, 31:161-175.

Phillips SJ, Dudík M, Schapire RE. 2004. A maximum entropy approach to species distribution modeling In: Proceedings of the 21st International Conference on Machine Learning, pp 655-662 New York: ACM Press.

Pimm SL, Jenkins CN, Abell R, Brooks TM, Gittleman JL, Joppa LN, Raven PH, Roberts CM, Sexton JO. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. Science 344(6187) 1246752.

Scolforo JR, Oliveira AD de, Carvalho LMT. 2008. Zoneamento ecológico-econômico do estado de minas gerais: Componente socioeconômico. Lavras: Editora da UFLA.

Scolforo JRS, Carvalho LMT. 2006. Mapeamento e inventário da flora nativa e dos reflorestamentos em Minas Gerais Lavras: Editora da UFLA 288 p.

Taylor N, Zappi DC. 2004. Cacti of Eastern Brazil. Royal Botanic Gardens, Kew, 499 p.

Taylor NP, Zappi D, Braun P, Machado M. 2013a. *Pereskia aculeata*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. (Downloaded on May 2015).

Taylor NP, Zappi D, Machado M, Braun P. 2013b. *Pereskia grandifolia*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. (Downloaded on May 2015).

Thomas, P. 2013. *Araucaria angustifolia*. The IUCN Red List of Threatened Species. Version 2014.3. <www.iucnredlist.org>. (Downloaded on May 2015).

Thuiller W, et al. 2005a. Climate change threats to plant diversity in Europe Proceedings of the National Academy of Sciences USA, 102:8245-8250.

Thuiller W, Richardson DM, Pysek P, Midgley GF, Hughes GO, Rouget M. 2005b. Niche-based modelling as a tool for predicting the global risk of alien plant invasions at a global scale. Global Change Biology 11:2234–2250.

Veloso HP, Rangel-Filho ALR, Lima JCA. 1991. Classificação da vegetação brasileira, adaptada a um sistema universal. IBGE, Departamento de Recursos Naturais e Estudos Ambientais, Rio de Janeiro.123p.

Wiley EO, McNyset KM, Peterson AT, Robins CR, Stewart AM. 2003. Niche modeling and geographic range predictions in the marine environment using a machine-learning algorithm. Oceanography 16:120–127.

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A, NCEAS Predicting Species Distributions Working Group. 2008. Effects of sample size on the performance of species distribution models. Divers. Distrib. 14: 763-773.

Zachos FE, Habel JC. 2011. Biodiversity Hotspots: Distribution and Protection of Conservation Priority Areas. Berlin Heidelberg: London New York Springer-Verlag.

Zuur AF, Leno EN, Smith GM. 2007. Analysing ecological data. Springer: New York

CHAPTER 5

SYNTHESIS AND RECOMMENDATIONS

5.1 Synthesis of key findings

Prior to the studies that make up this thesis, there have been a few studies that investigated drivers and associated factors of land use and cover changes in the Brazilian Atlantic Forest (Silva et al. (2007), Teixeira et al. (2009), Freitas et al. (2010), Lira et al. (2012), Freitas et al. (2013), Ferreira et al. (2015)). However, to date there has been no comprehensive study of relationships between metrics quantifying the deforestation and fragmentation of the Atlantic Forest in Minas Gerais, and variables quantifying the socio-economic and bio-geophysical context within which these processes are occurring. This has been the one of the aim of the work that makes up this thesis. I tested a relatively novel method of statistical analysis and applied a multi-scale approach in order to disentangle and elucidate these relationships. In addition, I modelled the potential distribution of threatened plant species to assess the effectiveness of existing protected areas in conserving these species. This thesis provides, therefore, an important advance in our understanding of deforestation and forest fragmentation drivers and the efficiency of protected areas to protect threatened species in the Brazilian Atlantic Forest, supporting management and conservation planning with valuable information.

The first two chapters of the thesis applied a recently developed machine-learning technique, random forest analysis (Breiman 2001) to investigate relationships between deforestation/fragmentation and socio-economic/bio-geophysical factors. In chapter 2, I selected an alternative from a wide range of possible approaches to provide an appropriate analytical comparator to RF: the classical stepwise multiple regression. I found that RF proved to be a promising methodology for identifying these relationships, and that it has the potential to be an effective tool for providing essential information for aiding land use management decisions. Building on the results of Chapter 2, I extended the application of RF using a multi-scale approach in Chapter 3, grouping municipalities at sub-regional, regional, and biome scales. Additionally, in Chapter 2, I was limited to using only those landscape metrics that presented normally distributed residuals from linear models, since I was using a classical approach as a comparator. However, one of the advantages of RF is that it does not assume any particular frequency distribution in its input variables (Cutler and Stevens 2006, Prasad et al. 2006). Thus, in Chapter 3, where I was using only RF, I had greater flexibility in selecting independent and dependent variables. The relationships were found to vary from place to place and across spatial scales, and some metrics were better explained by the RF models at the largest (biome) scale, while others were better explained at the smallest (sub-regional) scale. Although I have shown that there is no unique factor driving deforestation and forest fragmentation across all scales, I identified road density as the most common factor in the models explaining deforestation and fragmentation metrics variance all scales. This supports the findings of the study conducted by Freitas et al. (2010). In addition, some categories of factors occurred more commonly in the models than others. For example, factors describing the spatial distribution of the natural, agricultural and infrastructural elements of the landscape were more common, while those describing patterns of population, employment and legal institutions appeared to be less important to explain the variance of the metrics. This is somewhat at odds with views expressed in the literature that landscape patterns in areas populated by humans are strongly determined by socio-economic drivers (Lambin et al. 2001, Geist and Lambin 2002). This apparent contradiction may be resolved by the suggestion that the landscape patterns observed today may be governed by historical patterns of socio-economic activity (as well as bio-geophysical factors), that may not be reflected in the data on quasi-present day socio-economic conditions. Thus, there may be an asynchronous relationship between socio-economic drivers and patterns of deforestation and forest fragmentation.

Finally, in Chapter 4, I identified gaps in the protection of threatened plant species in three biomes in Minas Gerais (Atlantic Forest, Cerrado and Caatinga). To do so, I modelled species' potential geographical distributions using species occurrence points from a large dataset provided by the Vegetation Monitoring System Project. In total, 169 fragments of remaining natural vegetation scattered across the three biomes. Additional points were obtained from two other databases: the national Herbaria network (SpeciesLink) and NeoTropTree database. I then superimposed the resulting map of potential species distributions onto maps of existing natural vegetation remnants and protected areas. Thus, I identified those areas that are environmentally suitable for the threatened species that do not coincide with existing protected areas as gaps in protection, and, therefore, proposed them as priority areas for conservation. As expected, I found significant gaps in the existing network of protected areas, especially in the Atlantic Forest biome, which reinforces the need to prioritise the expansion of protected areas in this biome. This chapter also concluded that it is likely that species from the Cerrado and Caatinga biomes are missing from the red lists due to a lack of studies in these areas. Thus, conducting inventories to identify and locate threatened species in these biomes also requires prioritisation.

5.2 Challenges and limitations

Through the research reported in this thesis, I have examined the challenges involved in disentangling and elucidating the relationships between deforestation and forest fragmentation and socio-economic and bio-geophysical factors, and in assessing the gaps in conservation of threatened plant species in Minas Gerais. The inherent complexity and heterogeneity of a megadiverse biome such as the Atlantic Forest has probably contributed to the high level of unexplained variance in some of the models and affected the accuracy of some species distribution models

The main limitations and challenges I found conducting this thesis are as follows:

1. Despite using a huge dataset with a wide variety of independent variables, I still found a high level of unexplained variance in the RF models. This may be because important factors that drive deforestation and forest fragmentation are missing from the models. In particular, the relative lack of 'information on change over time' in the data implies that historical factors and asynchronous influences have not been identifiable by the modelling approach used here.
2. This lack of temporal information is one of the main caveats in this thesis. The deforestation and forest fragmentation metrics were derived from the vegetation monitoring system dataset (Scolforo and Carvalho 2006, Carvalho and Scolforo 2008, Carvalho and Scolforo – unpublished data), which comprises land cover maps from 2003 to 2011. Socio-economic and bio-geophysical variables were obtained from the ZEE-MG database, and the years for which these variables were collected were limited by the availability of information from national agencies, ranging from 2003 to 2006 (Chapters 2 and 3). This has undoubtedly limited exploration of the relationships between such variables.
3. Despite the fact that a set of multiple factors drive anthropogenic land use and cover changes such as deforestation and forest fragmentation, there are some specific factors which are critical in determining the explanation of metric variance, even when they

present a low percentage of variable importance in some models (Chapter 2 and 3).

4. I was limited to using landscape metrics that presented normally distributed residuals from linear models, since I was using a classical approach as a comparison method in Chapter 2, while in Chapter 3 I used only RF and therefore had a greater flexibility in choosing metrics. Thus, the metrics in Chapters 2 and 3 are not exactly the same, which limits comparisons between their findings.
5. The lack of strong evidence of associations between factors and metrics does not mean that they do not exist. This may be due to an asynchronous relationship between socio-economic drivers and patterns of deforestation and forest fragmentation. Additionally, the same way that the ecological consequences of anthropogenic disturbances may take a long time to become fully apparent in ecosystems (Hylander and Ehrlén 2013), the effects of socio-economic factors upon LUCC also may take a time to become apparent. Again, an examination across time would help to make this clear.
6. One limitation in Chapter 4 was that I had to match the IUCN list of threatened species with the list of species sampled in the vegetation monitoring system dataset. I used this dataset because samples were collected using a well-designed and systematic protocol covering a large number of fragments in the three biomes of Minas Gerais. However, as only higher plant species were sampled in the vegetation monitoring system, many small threatened species were removed from my final list of threatened species.
7. Not all areas indicated as suitable by the models will be occupied by the species (Anderson et al. 2003). This is because important factors

that determine the distribution of species are missing in the models, as are other abiotic factors, biotic interactions, scatter barriers, anthropogenic effects, populations extinction, stochastic events, and historical factors (Pearson and Dawson 2003, Soberón and Peterson 2005). Therefore, care should be taken, as it is still not possible currently to include in the models all factors that may restrict the species distribution.

5.3 Recommendations for Atlantic Forest management and conservation

The results from the three studies reported in this thesis combine to illustrate some important findings that can be used to assist the development of management strategies and conservation plans for the Brazilian Atlantic Forest at multiple spatial scales. In this respect, the most significant outcome of this thesis is the identification of the need for conservation strategies that take into consideration the potential drivers of deforestation and forest fragmentation in the Atlantic Forest. Although I demonstrate in this thesis that a complex set of factors affects deforestation and forest fragmentation, I also identify categories of factors that are more commonly associated with both deforestation and fragmentation across all of the spatial scales studied. Another important outcome from this thesis is the fact the relationships between metrics and drivers of deforestation and fragmentation vary from place to place and across spatial scales. This needs to be taken into consideration by planning conservation actions in a scale-appropriate manner: some approaches need to be adopted at a biome-wide scale, while others need more locally-adapted conservation actions.

Conservation planning should also account for the gaps in protection identified in this thesis. Considering the huge loss of original coverage of the three biomes of Minas Gerais, the current state of degradation of the Atlantic Forest and the Cerrado, and with the lack of knowledge in some regions, especially in the Caatinga, the utility of species distribution modelling (SDM) is emphasised by my findings. SDM can aid indication of environmentally suitable areas for reintroduction of species in biodiversity restoration projects using the maps created for the plant species studied in this thesis, which have been overexploited in the past.

It is noteworthy that the maps generated in this study are only one tool that can be used in biological conservation projects and other applications. There is also a pressing need for obtaining new field records of threatened species. These records will ensure the effective validation of SDM and the success of conservation strategies based on them.

5.4 Future research priorities

This thesis advances our knowledge on how socio-economic and bio-geophysical factors can interact with deforestation and forest fragmentation in the Brazilian Atlantic Forest and identifies the main gaps in protection of threatened species in three biomes of Minas Gerais. However, it also identifies the need for new studies that further expand our understanding of subjects such as the effects of multi-scale drivers upon land use change in this biome, so that its conclusions can be extrapolated to other tropical biomes. Future work should also investigate such relationships across different temporal scales, thereby addressing historical factors. Further research should also account for patterns and processes throughout the biome, considering variations among all regions where the Atlantic Forest occurs. Species distribution modelling should also be expanded to cover entire biomes. This may help support efforts to draw more general conclusions and extrapolate them to larger spatial scales and other tropical biomes. I specifically recommend the use of approaches that:

1. Include other socio-economic and bio-geophysical factors, which are missing from our dataset, in order to try to cover a full set of potential drivers;
2. Account for temporal investigations along with the multi-scale spatial approach introduced here, in order to capture associations between factors and metrics in more detail;
3. Expand the multi-scale spatial approach I have introduced here to other regions of the Atlantic Forest in order to better understand the effects of regional context. Understanding the factors that define this contextual specificity is very important for extrapolating the conclusions presented here to the rest of the Atlantic Forest, thereby supporting conservation planning; and
4. Expand species distribution modelling to cover the entire Atlantic Forest biome and thus identify biome-wide gaps in protection.
5. Model species potential distribution considering future scenarios of climatic change.

5.5 Concluding remarks

The Brazilian Atlantic Forest is an extremely heterogeneous and unique biome currently under severe threat. Conservation actions in this biome are needed more than ever to mitigate the consequences of potential threats and protect the biodiversity that remains. A few studies have been developed recently trying to address the potential threats and related factors in the Atlantic Forest. However, to date there has been no comprehensive study of relationships between metrics quantifying the deforestation and fragmentation of the Atlantic Forest, and variables quantifying the socio-economic and bio-geophysical context within which these processes are occurring. This thesis provides a comprehensive, quantitative and multi-scale assessment of such relationships, using a relatively novel statistical approach. It also provides strong evidence of the need for expansion of protected areas in the Atlantic Forest and other two biomes in Minas Gerais; based on gaps in protection identified by modelling the potential distribution of threatened plant species. This thesis expanded our understanding of related factors to the main threats to the Atlantic Forest in Minas Gerais. It also provided valuable information to support management and conservation planning in the Brazilian Atlantic Forest, Caatinga and Cerrado, assessing the efficiency of protected areas to protect threatened species.

While the findings presented by this research has contributed significantly to address the problems highlighted above, there are still many challenges to face and need for new studies, in order to extrapolated the thesis findings to other tropical biomes, as detailed in section 5.4 Future research priorities. Improved management and conservation strategies are urgent due to the rapid rates of biodiversity loss. In this sense, future work should extend the investigation on the relationships addressed here incorporating historical factors and different temporal scales. In addition, new studies should investigate variations among all regions, accounting for patterns and processes throughout the Atlantic Forest, as well as, to expand the species distribution modelling to cover entire biomes and to simulate future scenarios of climatic change.

REFERENCES

ANDERSON RP, LEWC D, PETERSON T. 2003 Evaluating predictive models of species'distributions: criteria for selecting optimal models **Ecological Modelling**, 162:211-232.

BREIMAN L. 2001. Random forests. **Mach Learn** 45:5–32.

CARVALHO LMT, SCOLFORO JRS. 2008. **Inventário Florestal de Minas Gerais**: Monitoramento da Flora Nativa 2005-2007. Lavras, Editora: UFLA, 357p.

CARVALHO LMT, SCOLFORO JRS. Unpublished data. **Inventário Florestal de Minas Gerais**: Monitoramento da Flora Nativa 2007-2009; 2009-2011.

CUTLER A, STEVENS J. 2006. Random Forests for Microarrays. In: Kimmel A, Oliver B (Eds). **Methods in Enzymology.** Academic Press, San Diego, pp 422–432.

FERREIRA MP, ALVES DS, SHIMABUKURO YE. 2015. Forest dynamics and land-use transitions in the Brazilian Atlantic Forest: the case of sugarcane expansion. **Regional Environmental Change** 15:365-377.

FREITAS MWD DE, SANTOS JR DOS, ALVES DS. 2013. Land-use and land-cover change processes in the Upper Uruguay Basin: linking environmental and socioeconomic variables.**Landsc Ecol** 28:311-327.

FREITAS SR, HAWBAKER TJ, METZGER JP. 2010. Effects of roads, topography, and land use on forest cover dynamics in the Brazilian Atlantic Forest.**Forest Ecology and Management** 259:410-417.

GEIST HJ, LAMBIN EF. 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. **BioScience**. 52 (2): 143-150.

HYLANDER K, EHRLÉN J. 2013. The mechanisms causing extinction debts.**Trends in ecology and evolution** 28:341–6.

LAMBIN EF. et al. 2001. The causes of land-use and land-cover change: moving beyond the myths.**Global Environmental Change** 11:261–269.

LIRA PK.et al. 2012. Land-use and land-cover change in Atlantic Forest landscapes.**For Ecol Manage** 278:80–89.

PEARSON RG, DAWSON TP. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? **Global Ecology and Biogeography** 12:361-371.

PRASAD AM, IVERSON LR, LIAW A. 2006. Newer Tree Classification and Techniques: Forests Random Prediction Bagging for Ecological Regression.**Ecosystems** 9:181–199.

SCOLFORO JR, OLIVEIRA AD DE, CARVALHO LMT. 2008. **Zoneamento ecológico-econômico do estado de minas gerais**: Componente socioeconômico. Lavras: Editora da UFLA.

SCOLFORO JRS, CARVALHO LMT. 2006. **Mapeamento e inventário da flora nativa e dos reflorestamentos em Minas Gerais.** Lavras: Editora da UFLA 288p.

SILVA WG.et al. 2007. Relief influence on the spatial distribution of the Atlantic Forest cover on the Ibiúna Plateau, SP. **Braz J Biol** 67(3):403-411.

SOBERÓN J, PETERSON AT. 2005. Interpretation of Models of fundamental ecological niches and species' distributional areas. **Biodiversity Informatics** 2:1-10.

TEIXEIRA AMG.et al. 2009. Modeling landscape dynamics in an Atlantic Rainforest region: Implications for conservation. **For Ecol Manage** 257:1219–1230.