Modelling Land Cover Change in Tropical Rainforests

Isabel Maria Duarte Rosa

Imperial College London

Division of Ecology & Evolution: Department of Life Sciences

Silwood Park Campus, Ascot, Berkshire, SL5 7PY, U.K.

A thesis submitted for the degree of

Doctor of Philosophy

Diploma of Imperial College (DIC)

September 2013

Supervisors: Dr. Robert Ewers

Dr. Drew Purves

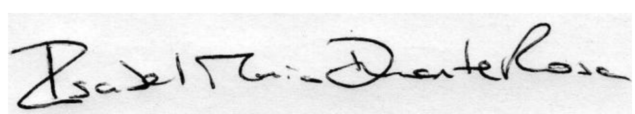
Para o João, por estar sempre ao meu lado

To João, for always being there for me

Declaration of Originality

I hereby certify that all content of this thesis is my original research and collaborations with others researchers are fully acknowledged. I am senior author in all Chapters, Dr. Robert Ewers co-authored all data Chapters (Chapters 2 to 6) which reflect his strong input in the development of the concepts, advice on the methods and support in the editorial process of all manuscripts. Dr. Drew Purves co-authored Chapters 4 to 6 and helped with the C++ programming of the model developed in this thesis. Other co-authors are Dr. Carlos Souza Jr. (Chapters 2 and 4), Sadia Ahmed (Chapter 3), Dr. Matthew Smith (Chapter 6), and Dr. João Carreiras (Chapter 5). All of them provided data and/or invaluable comments to further improve the manuscripts. In Chapters 2, 4, 5 and 6 I collected and analysed all the data utilised, made all figures showing the results, wrote the first draft of the manuscript and edited all subsequent drafts of the manuscripts. In Chapter 3, Sadia Ahmed and I submitted the manuscript as joint first-authors, since we both collected the data and collaborated throughout the whole manuscript and review process.

20th September 2013



(Isabel Maria Duarte Rosa)

Copyright declaration

Acknowledgements

First and foremost I owe profound thanks to my Imperial College supervisor, Dr. Robert Ewers. Throughout my PhD Rob has been very patient and has given me his constant support and guidance. The openness to share his knowledge made me learn immensely in our meetings and his constant encouragement and peculiar sense of humour made me truly enjoy working with him.

I am very grateful to my Microsoft Research supervisor, Dr. Drew Purves, for his strong enthusiasm always shown in our meetings and crucial input for the success of my PhD.

The work I have done in this thesis would not have been possible without my co-authors. Therefore, I would like to thank Dr. Carlos Souza Jr. from Imazon, Dr. Matthew Smith from Microsoft Research and Dr. João Carreiras from Tropical Research Institute, for providing data, ideas and comments, which were very much appreciated. Special thanks go to Sadia Ahmed for her friendship, our lively discussions, and the many good moments we have shared during these last few years.

For funding all the research done in this thesis I would like to thank Microsoft Research Cambridge and the Grantham Institute for Climate Change. For providing access and help with the High Performance Computer at South Kensington I would like to thank Simon Burbidge, Will Pearse and James Ronsindell. In addition, I also thank Diana Anderson, Christine Short and Paola Riboldi-Gomm for all the support with logistics and administrative work.

I am also deeply grateful to many people who have supported me throughout this challenging period of my life. For making my time at Silwood Park so enjoyable I would like to thank all my colleagues and friends at the Forest Ecology and Conservation research group of which I am a part: Marion Pfeiffer, Veronique Lefebvre, Oriana Pilia, Poppy Lakeman-Fraser, Charlie Marsh, Cristina Banks-Leite, Ollie Wearn, Hayley Brant and Andrew Bradley. I would also like to show my gratitude to many of my colleagues and friends back in Portugal who always showed their support and friendship throughout the years: Ricardo Lourenço, João Barreiros, Alzira Ramos, Ana Barros, Ana Sá, Joana Melo, Sofia Oliveira, Duarte Oom, João Silva, José

Miguel Cardoso Pereira e Cristina Gabriel. I would like to say a very special thanks to my friend, Andreia Silva, for all these years of friendship, with many long and exciting conversations and for her constant support.

I am truly thankful to my family, in particular my sister for being an amazing big sister; to my parents for showing their constant encouragement and support; and to my grandmothers who were always an example for me. Finally, I want to thank a very special person, João, for always being there for me.

Abstract

Tropical deforestation is one of the most important drivers of biodiversity loss and carbon emissions. This thesis seeks to analyse the dynamics of tropical deforestation and develop a probabilistic model that predicts land cover change (LCC) in the tropics. The main findings from the analysis of the Brazilian Amazon deforestation dynamics are that large clearings comprised progressively smaller amounts of total annual deforestation while the number of smaller clearings remained unchanged over time. These changes were coincident with the implementation of conservation policies by the government.

The review of LCC models presented here showed that this modelling community would benefit from improving: the openness to share model inputs, code and outputs; model validations; and standardised frameworks to be used for model comparisons. The modelling framework developed aimed to tackle the limitations found before and two scenarios of deforestation in the Brazilian Amazon were simulated. For both scenarios forest next to roads and areas already deforested were found to be more likely to be deforested. States in the south and east of the region showed high predicted probability of losing nearly all forest outside of protected areas by 2050.

The release of carbon to the atmosphere is an important consequence of tropical deforestation. Even if deforestation had ended in 2010 there would still be large quantities of carbon to be released. The amount of carbon released immediately is higher than the one committed for future release in the first few years of analysis, but presently these accounted for at least two-thirds of total carbon emissions. Finally, the drivers of LCC were found to vary among transition types, but less so through time. The accuracy of the model predictions was heavily dependent on the year calibrated, suggesting that a widespread reliance on single calibration time period may be providing biased predictions of future LCC.

Table of Contents

List of tables

List of figures

List of appendices

Chapter 1 - Introduction

The importance of tropical forests as an ecosystem has long been recognised; however the fast rate of destruction that these forests have been facing over the last decades resulted in a significant reduction of their area, and places them in the danger of reducing their extent even more significantly in the overcoming decades. In this thesis, the importance of tropical forests in terms of ecosystem services and influence on climate, carbon and water cycles is discussed at the global scale. Then, the causes and consequences of the threats that these forests face are examined at regional scales (Amazon, Central Africa and Southeast Asia), with particular focus on the Brazilian Amazon. Not only this is the region where the highest amount of forest area is lost every year, but Brazil is also the country where tropical deforestation is best monitored and, as a consequence, best studied and understood.

The main aim of this thesis is to develop a model to predict the location and the amount of destruction in tropical forests. Over the next chapters the structure of the model will be presented and I will analyse its performance in different regions and at different spatial scales. I will conduct a detailed examination of the model's sensitivity to calibration data and its ability to replicate patterns of forest abandonment and regeneration as well as direct deforestation. Finally, I will use the predictions from the model to work out what the trends in historical destruction across the world's wet tropics means in terms of carbon emitted to the atmosphere.

1) The importance of tropical rainforests

Tropical rainforests (Fig. 1) cover only as much as 10% of the Earth's land surface, but they still represent the largest forest extent on Earth (Hansen et al. 2010), and they are believed to be home to about half of the world's biodiversity (Wilson 1988). Tropical rainforests are naturally very dense with multiple layers of continuous vegetation which is a consequence of the unique environmental conditions found in the tropics: contrarily to what happens in temperate regions, there is no strong seasonality of temperature, temperatures are warm and average roughly 25°C (Chambers et al. 2000; Araújo et al. 2002; Vourlitis et al. 2002) and there are high levels of precipitation (~ 2,000 mm) that can be evenly distributed throughout the year (Borchert 1998; Chambers et al. 2000; Vourlitis et al. 2002). These conditions create an ideal environment for forest growth.

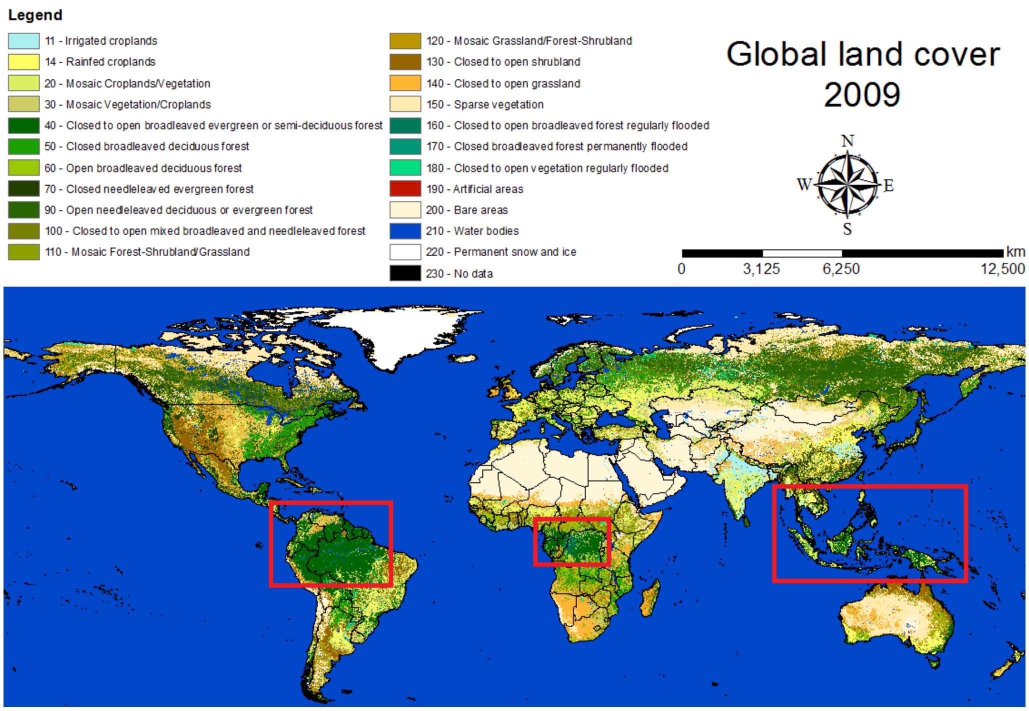


Figure 1 – Global land cover in 2009, with tropical regions indicated by red squares (data source: GlobCover, European Space Agency).

These forests provide a variety of important ecosystem services, both at a local and global scale. A number of indigenous communities live in tropical forests (Zimmerman et al. 2001), and they benefit from ecosystem services provided by them such as bushmeat (Fa et al. 2002). Tropical forests contribute to the regulation of water flows in rivers (Costa et al. 2003), and affect climate not only at regional scale (Nobre et al. 1991), but also at global scale (McGuffie et al. 1995). Their important role in the carbon cycle as carbon sinks has long been recognized (Phillips et al. 1998; Malhi & Grace 2000; Lewis et al. 2009b). However, during the last decades tropical forests have been facing severe destruction and degradation related to human activities.

Losing such biodiversity-rich ecosystems at such a fast rate is a major threat to the world's biodiversity (Myers 1988). In addition, the fast destruction of tropical forests is compromising the future of many indigenous people (Alcorn 1993) as well as the future of local populations who use forests as source of food, construction materials, remedies, and also sell forests products at local markets to get their family income (Laurance 1999). The destruction of tropical rainforests will also eliminate the important role these forests play in stabilizing soils and rivers, preventing massive soil erosion, soil compaction and floods, as well as alter the regional and global climate (Foley et al. 2003). In the Amazon, rainfall is closely linked to the existence of forests, with about half of it originating from plant evapotranspiration (Salati & Vose 1984), which means that destroying these forests would reduce the air humidity, increase the surface temperature and the dry season severity, leading to a much higher probability of wildfires occurrence which would have devastating impacts in the landscape (Nepstad et al. 1998). This phenomenon observed at regional scale can also happen at large-scale. Clearing tropical forests extensively may create a warmer, drier climate (Costa & Foley 2000; Cox et al. 2000; Cox et al. 2004).

Moreover, emissions from land use change are now the second-largest anthropogenic source of carbon dioxide (CO_2), just behind fossil fuel emissions (Le Quere et al. 2009). Houghton (2005) estimated that tropical deforestation and degradation account for about 20-25% of the annual total anthropogenic emissions of greenhouse gases. However, in the global carbon budget, carbon emission estimates from land-use change have the lowest precision associated. This is due to the uncertainties in the true deforestation rates (Grainger 2008), the amount of biomass and soil carbon held within different forest types (Keith et al. 2009), as well as the spatial distribution of forests (Achard et al. 2004).

2) Threats to tropical rainforests

Geist and Lambin (2002) identified two types of causes for tropical deforestation: proximate causes (*e.g.* infrastructure expansion, wood harvesting and agriculture expansion) are human activities that directly lead to change at the local level; and underlying causes, which can be demographic (population dynamics), economic (economic growth or change), technological (improvement or progress) or political (environmental laws or policies). In line with previous findings, Gibbs et al. (2010) stated that in the 1980s and 1990s, rainforests (intact or disturbed) were the main source (80%) of new agricultural land, especially in countries like

Brazil or Indonesia. This means that although the world was successful in increasing food production, that increase was obtained at the expense of tropical forests (Foley et al. 2005). Tropical deforestation is a dynamic phenomenon among regions: Southeast Asia has the highest rates of deforestation and Central Africa the lowest. However, although there has been a significant reduction in deforestation rates in the last few years in the Amazon Basin (Nepstad et al. 2009), it is still the region where people destroy the highest absolute amount of forest area each year (Rudel 2005).

2.1) Deforestation in the Amazon

The Amazon is the largest remaining continuous tropical rainforest on Earth. It covers about 6 million km^2, crossing nine nations' boundaries. Brazil is the country that hosts the largest portion (about 60% of the area) of the Amazon. This region is characterised by its high cultural and biological diversity (Barlow et al. 2011). However, by 2009 already 15% of the forest cover in the Legal Amazon had been deforested (Pereira et al. 2010a).

In the late 1960s and early 1970s South American governments started to end the passive protection of the Amazon. Until then, there were no roads to access the forest and only rivers were used to transport wood from the forest sites to the markets in the city centres, which make it a very costly economic activity. This end of protection came in the form of colonisation schemes which included expansion of the under-developed road networks as well as land settlement schemes along these new highways (Steininger et al. 2001; Fearnside 2005; Etter et al. 2006b). The aim of these programs was to take poor and unemployed people from the big cities and grant them a portion of land in the Amazon so they could settle there and live from forest and agriculture related activities. However, many of these people did not have the knowledge to be successful farmers, and many of the smallholders who worked to establish claims to land at the forests' margins did not have enough capital to establish commercially successful farms. Therefore, after clearing an appreciable portion of their land, they would often sell out to larger and better-financed landowners (Diegues 1992). This migration of people was particularly strong in the Legal Amazon where in about forty years, the population rose sharply from about 8 million people in 1970 to 24 million in 2009 (Pereira et al. 2010a).

Particular focus will be put on the Brazilian Amazon because this is the region where most of the research presented in this thesis was performed. Further, Brazil is the tropical nation with the strongest deforestation monitoring program, maintained by the Instituto Nacional de Pesquisas Espaciais (INPE) for more than 20 years.

2.1.1) Temporal dynamics of deforestation in the Brazilian Amazon

In Brazil there are two concepts of Amazon: there is the Amazon biome and the Legal Amazon. The first gets the name after the typical forest cover found in this region and it represents around 4 million km^2 (49% of the Brazilian territory). The second, in addition to the area covered by the Amazon biome – mainly dominated by tropical forests (63%), and some areas of savannah-like vegetation called 'cerrado' (22%) – covers about 5 million km^2 (59% of the Brazilian territory) (Pereira et al. 2010a). Nine Brazilian states (Acre, Amapá, Amazonas, Maranhao, Mato Grosso, Pará, Rondônia, Roraima and Tocantins) are included in the Legal Amazon, although Maranhão and Tocantins are only partially included.

The rate of deforestation in the Brazilian Amazon (Fig. 2) is highly dynamic (Ewers et al. 2008). These fluctuations are related to several factors such as the economic health of the country, infrastructure development, and the world's demand for agriculture products, such as beef or, more recently, soybeans (Fearnside 2005; Morton et al. 2006; Nepstad et al. 2006b; Ewers et al. 2008). Although these factors drive the deforestation rates in the Amazon, the intensity of the different processes vary greatly across the region due to regional differences, including physiographic attributes, access to infrastructures, population characteristics and dynamics and socioeconomic organization (Garcia et al. 2007).

Brazil's economic recession best explains a decline in deforestation rates observed from 1987 through 1991 (Fearnside 2005). Ranchers were unable to expand their clearings as quickly, and the government lacked funds for road construction and settlement projects. This was reversed in 1995 with a peak in deforestation rates that was probably a reflection of economic recovery under the Real Plan ('Plano Real', a set of measures taken by the government to stabilize the Brazilian economy in early 1994). The reforms increased the availability of capital, and municipal elections in 1994 resulted in an increase in agricultural credit. The association of major swings in deforestation rate with macroeconomic factors such as money availability and inflation rate is one indication that much of the clearing is done by those who invest in medium to large cattle ranches, rather than by small farmers who use family labour (Fearnside 2005).

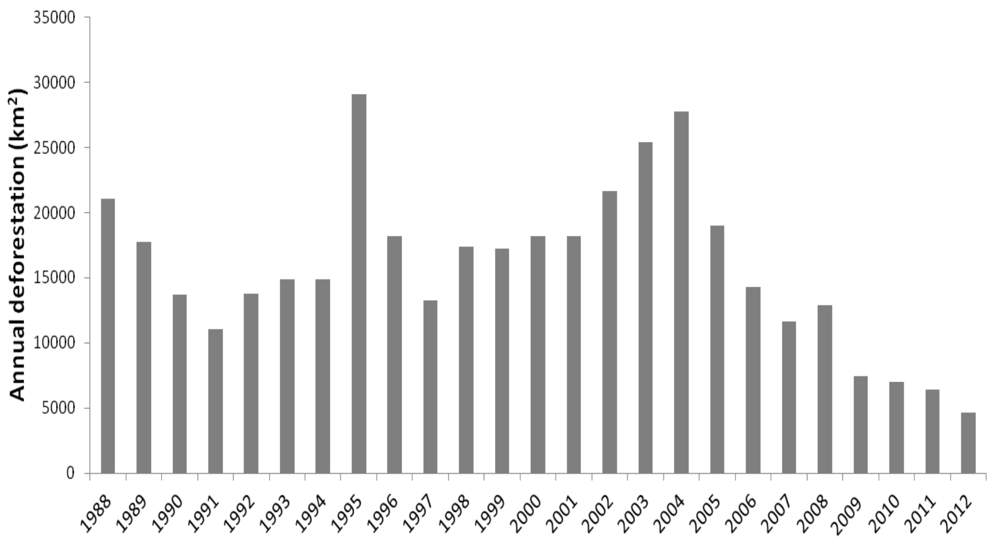


Figure 2 – The annual rate of deforestation in the Brazilian Amazon (data source: INPE).

Cattle ranching is now the most common land use in deforested areas throughout the Brazilian Amazon (Kaimowitz et al. 2004). However, this region did not even produce enough beef to feed its own population until 1991. After that, the region began to produce a growing surplus of beef for the national market, but Brazilian beef exports to other countries continued to play a minor role. As recently as 1995, Brazil exported less than 500 million US dollars of beef. By 2003, only eight years later, Brazil was exporting three times as much, 1.5 billion US dollars (Kaimowitz et al. 2004). Last decade's immense expansion of cattle ranching is explained by the eradication of foot-and-mouth disease, as well as the devaluation of the Brazilian currency, the outbreaks of bovine spongiform encephalopathy that took place in Europe, and also to improvements in beef production systems (Nepstad et al. 2006b). Recently, the planting of intensive annual crops, such as soybeans, in areas formerly used as pastures is displacing cattle ranching from the margins to the core of the region (Laurance 2007; Barona et al. 2010). The expansion of soybean production in recent years is especially related to the development of soy varieties tolerant to the Amazon climate, the growing worldwide demand for soybeans for animal feeding or biofuels, which contributed to the improvement in storage and processing facilities in this region (Nepstad et al. 2006b), and changes in the crops grown in other nations due to government subsidies (Laurance 2007).

After 2004 deforestation fell substantially which is believed to be related with the global economic recession (Nepstad et al. 2009), which lead to a drop in the soy prices, as well as with successful conservation (expansion of protected areas network and market based strategies) and law enforcement (control and command field operations) policies applied by the Brazilian government, which also affected the spatial configuration of deforestation in the region (Rosa et al. 2012).

2.1.2) Spatial dynamics of deforestation in the Brazilian Amazon

In absolute terms, Mato Grosso and Pará are the states that register the highest amounts of deforestation each year. They are included in the so-called Arc of Deforestation, where most of the deforestation occurs, along with the Southeast of Maranhão, the North of Tocantins, Rondônia, and the Southeast of Acre (Fig. 3). The concentration of deforestation in this region of the Amazon is related to the proximity to urban centres (markets) and roads, reinforced by the higher connectivity to the more developed parts of Brazil, and the more appropriate climatic conditions for agriculture in comparison to the rest of the region (Aguiar et al. 2007).

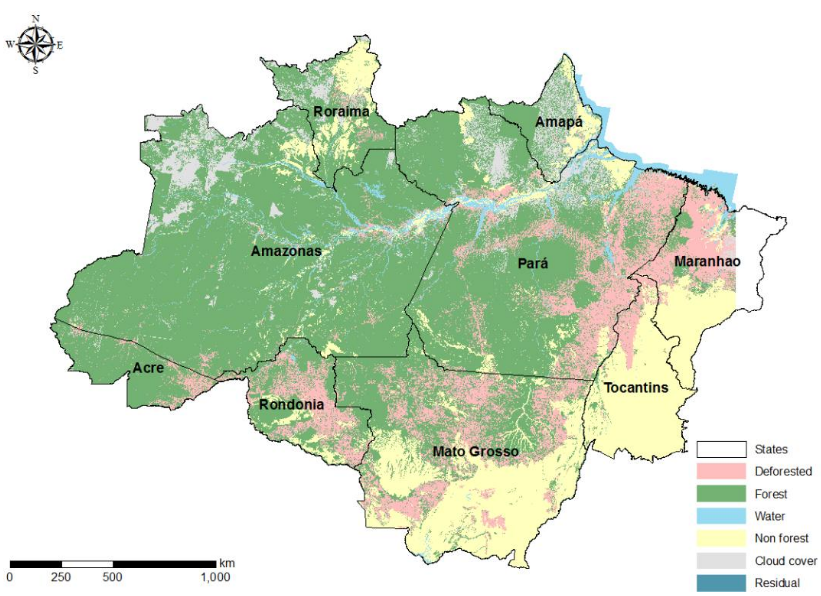


Figure 3 – Land cover in the Brazilian Amazon by 2011 (data source: INPE).

Climate and soils are the main constraints to agriculture (Sombroek 2001). The inaccessibility and humid climate of the more western-northern regions of the Amazon make states like Amazonas or Roraima less agriculture-prone (Fig. 3). On the other hand, along the Arc of Deforestation, easy access to forests and adequate road networks, which make forest products transportation easy and cheaper, explains the concentration of large deforestation occurring in this particular region of the Amazon (Chomitz & Thomas 2003). Large-scale clearings are almost always attributed to the expansion of cattle ranching, which is sometimes displaced and pushed further into the forest by the soybean industry (Ewers & Laurance 2006). On the other hand, small-scale subsistence farming dominates old occupation areas such as the Northeast of Pará and Maranhão, and also in some municipalities in the North of Mato Grosso (Aguiar et al. 2007).

Mato Grosso is in the frontline of large forest clearings, followed by Pará and Rondônia (Alves 2002; Aguiar et al. 2007; Pacheco 2009), which in the Brazilian Amazon are usually associated with cattle ranching and soybean plantations (Chomitz & Thomas 2001; Lorena & Lambin 2009). In the Brazilian Amazon, Mato Grosso is the state with the largest amount of soybean plantation area (87% of the total area), as well as the state with the largest amount of pastureland (35% of the total area) and temporary agriculture (42% of the total area) (IBGE 2006). Pará and Rondônia have the largest amount of permanent agriculture (40 and 33% of the total area respectively), and also significant amounts of pasture (24 and 16% of the total area respectively).

The most important driver of deforestation in the Amazon is, however, roads because they provide easy access to forests and a fast link between these and the markets. There are two types of roads in the Brazilian Amazon: official and unofficial (Brandão & Souza 2006). The official roads are built by the government whereas the unofficial roads are built by farmers and logging companies (Perz et al. 2007). They are useful for rural populations by reducing their isolation; however, they have important environmental impacts. When a new road is constructed, the forest cover is reduced progressively, due to the slash-and-burn agriculture practice by small farmers who settle along these roads. As the land loses its fertility, usually after 2 to 5 years, they move deeper into the forest clearing small portions of land until there is almost no forest at all. Of special concern are unofficial roads – being built without the planning and authorisation required by law (Brandão & Souza 2006).

Protected areas have been established in the Amazon to avoid deforestation in high-pressure areas. Especially, in the Brazilian Amazon these cover a significant share of the remaining forest cover. Recent studies show that federal conservation lands and indigenous reserves both suffered far fewer impacts of logging and deforestation than unprotected lands (Nepstad et al. 2006a; Soares-Filho et al. 2010). Deforestation in protected areas is illegal and, overall, human pressure in protected areas is much smaller than in unprotected areas. However, this pattern might change in the near future. In Rondônia, a state with 45% of its area considered protected, but where a third of its forest cover is already deforested, deforestation in 2004 affected 6.3% of the protected areas (the average for the Amazon is only 1.7%)(Ribeiro et al. 2006).

2.2) Deforestation in Central Africa

The second largest rainforest in the world is located in Central Africa. It covers an area of around 2 million km^2 and is spread along six countries: Cameroon, Democratic Republic of Congo (DRC), Republic of Congo, Gabon, Equatorial Guinea and Central African Republic. DRC is, however, the nation which contains the largest share (63%) of the forest (Laporte et al. 2007), often referred to as the Congo Basin. About 85 million people live in Central Africa and with a strong population growth (2.8% per year) this number is expected to grow largely over the next decades (de Wasseige et al. 2012). Most of these people harvest forest products for both food and domestic energy and depend on slash-and-burn shifting agriculture for subsistence. Contrary to what happens for the Amazon, there is a huge lack of information on the actual forest extent, deforestation and degradation in the Congo Basin. This is the most challenging region to monitor due to the persistence of cloud cover and the fine scale of forest cover change dynamics (Duveiller et al. 2008). Nonetheless, it is believed that this is the tropical region with the lowest deforestation rates (although estimates vary greatly), which can be explained by the lack of a good infrastructures and markets for wood products.

Prior to the 1990s, deforestation was mainly driven by logging and small-scale subsistence farming, however, it is hard to understand the extent of the colonisation impacts in the central African forests due to lack of systematic data (Wilkie & Laporte 2001). The inaccessibility of Congo Basin forests has prevented the widespread exploitation of this region, but in recent decades human-related activities (*e.g.* mineral exploration) have steadily increase their impact on the land-use dynamics (Rudel et al. 2009). Deforestation in this region is much more pronounced closer to the urban areas, where small farmers grow food crops to sell on the urban markets. As the cities and the road networks around them grow in size, the amount of land cleared around those also increase. New roads decrease the costs of transporting goods to markets and make it profitable to continue cultivating land (Rudel 2005). In addition, during periods of political stability people settle close to the rivers and roads, opening corridors of cleared land within the forests.

In the 1990s the logging companies increased their activity and started developing a better road network, however, the existence of relatively well paid jobs in the cities and the long distances between the forests and the cities prevented the number of small farmers wanting to settle along these new roads to grow large (Rudel 2005). The reduced number of people living in more remote areas does not have the organizational capacity to lobby for new roads to be built by the government. Nevertheless, since the resources available are immense, these people usually practice a much more extensive form of shifting cultivation than the permanent agriculture practiced by farmers near the cities (Rudel 2005).

Between 1990 and 2000, deforestation in the Congo Basin averaged 0.21–0.42% per year (Zhang et al. 2005; Duveiller et al. 2008), with DRC and Cameroon reporting the highest rates, and Gabon reporting the lowest. Recent studies (Laporte et al. 2007; Duveiller et al. 2008) increase the concern on the future of these forests as they report an active deforestation frontier in the DRC where most of the remaining tropical forest is located. The improvement of the road networks by the logging companies (Laporte et al. 2007) and the presence of the Congo River ease the access to forests. Similarly to what happen in the Amazon, there are, however, good indications that protected areas are being effective in preventing deforestation (Duveiller et al. 2008).

Although deforestation is relatively low in this region, growing population densities and increasing allocation of logging concessions are expected to lead to higher deforestation rates due to the increase in demand for agriculture and forest products (Zhang et al. 2005). As international companies exhaust the heavily exploited deposits of oil, diamonds and copper in the region, the pressure on forests is expected to grow in the coming decades (Rudel 2005). Deforestation in Central Africa is still mainly due to logging, which in this region is still very selective and directed towards the most-valuable timber species, causing less impact on forests than in other tropical regions. Cameroon and Gabon are the countries with highest timber production, while DRC reports the lowest timber production. Most of the harvested timber is exported (unprocessed logs represent the largest exported product from Central Africa) rather than processed in the region (Duveiller et al. 2008).

2.3) Deforestation in Southeast Asia

Southeast (SE) Asia contains the world's third largest tropical forest distributed across countries including Brunei, Indonesia, Laos, Malaysia, Papua New Guinea, Philippines, Singapore, Thailand, Timor Leste and Vietnam. This region is home to about 250 million people and has experienced rapid development in the last decades. Every year, SE Asia reports the highest deforestation rates (1.2% per year), which keep causing rapid declines in the forest area (Miettinen et al. 2011). The exploration of rubber and oil palm (Abdullah & Hezri 2008) as well as the need to feed rapid growing population densities (Sasaki et al. 2009) motivated the conversion of forests to agricultural cultivations, which caused the majority of the forest damage in the countries of SE Asia. The revenues from these activities helped the economic growth of these countries leading to expansion of urbanization which also caused great damage in the local forests (Abdullah & Nakagoshi 2007). The decline in forest area is mainly due to three types of land cover change: forest degradation due to intensive logging, forest converted to large-scale agricultural plantations, and the expansion of small-scale farming (Miettinen et al. 2011).

The history of deforestation in this region is very similar to the one described above for the Amazon. Prior to the 1990s, government colonization programs in developing countries such as Indonesia were responsible for building roads which ended the passive protection of the forests, opening access to them and facilitating the transportation of forest products to markets (Rudel 2005). Rural populations increased greatly during this period, as these programs resettled 5 million migrants in Indonesia, who establish themselves along the roads opened by logging companies. The fact that the logging companies were there first, removing the trees, reduced the effort of the migrants in the creation of new farms and in preparing the land for cultivation (Chomitz & Griffiths 1996). In the Philippines, the rapid decline of forest cover between 1960 and 1987 lead to a six-fold increase of the area devoted to agriculture (Verburg & Veldkamp 2004).

After 1990, the pattern changed, with agribusinesses becoming much more important in explaining deforestation rates and leaving the forest heavily fragmented (Rudel et al. 2009). Nowadays, logging is only seen as a stage in the exploitation of land, with the main purpose of clearing forests in SE Asia being to expand farm for agriculture production such as oil palm and coffee plantations. Low labour costs, high oil palm prices and vast resources of land lead to an amazingly fast expansion of oil palm plantations. By 1998 Indonesia had around 3 million ha of oil palm plantations, becoming the world's second largest exporter of oil palm (the first was Malaysia) (Colchester et al. 2006). The area increased steadily during the last decades and by 2005 Indonesia had around 5 million ha of oil palm plantations (Colchester et al. 2006). The human influence in the forests is so great that the more accessible forests (in the lowlands, near roads and coastal areas) are believed to have been logged at least once (Rudel 2005). Only the accentuated topography of some islands has prevented the forest destruction to be even greater. However, growing international demand for tropical hardwoods as well as oil palm, which lead to planned expansion of oil palm plantations until 2020 (Colchester et al. 2006), increase the preoccupation on the future of the forests in this region.

3) Modelling land-cover change

There is a growing concern about the future of tropical rainforests. Predicting not only the amount of forest that will be lost in the future, but also the location of this loss is vital for a successful intervention of conservation planners and environment managers (Mertens & Lambin 1997). Tropical deforestation is probably the most paradigmatic example of land-cover change and how it can have deterministic impacts on the future of the planet and on human well-being. Modelling anthropogenic processes is still a great challenge. Not only is the physical environment itself highly variable from one region to another and constantly changing, but the underlying processes that drive change are usually very complex, combining socio-economic, cultural, political and environmental factors. To understand the complex interactions among human and biophysical factors that drive deforestation, numerous models of tropical deforestation have been developed (Hall et al. 1995; Mertens & Lambin 1997; Laurance et al. 2001b; Soares-Filho et al. 2002; Verburg et al. 2002). These models share common features such as the use of geo-referenced data, but what stands out is the amazing variety of methodologies and approached used in this field.

Spatial models of land cover change are developed to address a range of questions, including why changes happened in the past and identifying the main drivers of change; predicting how much and where the changes will occur in the future; and testing possible scenarios of landscape evolution and elucidating their biological, ecological and climatic consequences. One important aspect is the scale at which the model is developed and applied. There is no rule on the scale that should be used; instead this should be a compromise between computational power and the spatial pattern of the phenomenon being modelled. Further, the aim of the model can also help to decide the appropriate scale. Coarse scale models are mainly used to identify the overall patterns of land cover change and to define 'hot spots' of change. Finer scale models can provide more detailed insights in the pattern of land cover change as well as its consequences for ecological processes (Verburg and Veldkamp 2004).

Specifically for the Amazon, there have been many attempts to model land use / land cover change using very different approaches. Since examining these variations is the aim of the review presented in the Chapter 3 of this thesis, I refer the reader to that section for further details. For the other tropical regions such as Central Africa and SE Asia the number of regional spatially-explicit predictive deforestation models is low. Most of the predictive models found for these regions are an outcome of global land-cover/land-use change models (Feddema et al.

2005) or local scale models (Mertens & Lambin 1997; Prasetyo et al. 2011). This is not so surprising when we think of the lack of systematic deforestation monitoring data for these regions, especially for the Congo Basin. Global models are too generic (10– 50 km^2 pixel size) to assess accurately the areas more likely to be deforested within a region while local scale models are too site-specific to be applicable in other regions or "scale-up" to regional level.

In the Congo basin, Mertens and Lambin (1997) used an empirical model to examine the deforestation spatial patterns in southern Cameroon. Using a logistic regression with deforestation rate as predicted variable and proximity to roads, proximity to towns, proximity to the forest edge, and forest fragmentation as explanatory variables, the authors assessed the areas more likely to be deforested in the future. For SE Asia there are still some examples of land cover change models, especially for Indonesian islands. Prasetyo et al. (2011), for example, predicted the future deforestation at a 10 km scale in the Java Island under two population growth based scenarios. The authors used elevation, roads, slope, previous deforestation and population density as predictors. Kinnaird et al. (2003), on the other hand, calculated the probability of losing forest as a function of elevation and slope in the island of Sumatra and extrapolated that to predict future deforestation in the island. Also for Sumatra, Gaveau et al. (2009a) simulated deforestation from 2006 until 2030 at a 90 m scale, under three scenarios, using a logistic regression. Authors found that roads and previous deforestation were the main predictors of changes, whereas slope and protected areas slowed deforestation and the effect of population was not significant. Verburg and Veldkamp (2004) developed a model to identify overall patterns of land use change and 'hot zones' of deforestation in two regions of the Philippines, under two different scenarios (baseline and governance). The rate is calculated separately, by extrapolating past trends and then allocated using a spatial model based on highest suitability for change. This suitability is calculated using a logistic regression where the occurrence of deforestation is linked to the drivers, such as roads and population.

4) Tropical deforestation and the global carbon cycle

Anthropogenic causes, in particular, fossil fuel burning emissions from industrial development and global agricultural expansion are behind the increasingly rising concentration of CO_2 in the atmosphere (Le Quere et al. 2009). These effects were accelerated since the industrial revolution and by 1990 already 20 to 30% of the original forest extent had been lost, contributing 45% to the increase of atmospheric CO_2 observed since 1850 (Malhi et al. 2002). Fortunately, part of these emissions are dissolved in the ocean or taken from the atmosphere by terrestrial ecosystems (Prentice et al. 2001). In forest ecosystems, carbon is accumulated through the absorption of CO_2 and its assimilation into biomass (50% of it is carbon) and then stored in various pools such as above and below ground living biomass, timber, branches, foliage, roots, litter, woody debris, soil organic matter and forest products (Malhi et al. 2002).

Forest ecosystems can also act as source of carbon when land-use/cover change takes place, causing plant removal and affecting the organic content of soils. In addition, land clearing can lead to soil degradation and erosion, which can reduce the ability of the ecosystem to act as a carbon sink (Taylor and Lloyd, 1992). Currently these systems are acting as sinks at the global scale (Prentice et al. 2001), even with high quantities of carbon being released to the atmosphere due to deforestation, especially in the tropical regions (Houghton et al. 2005). However, if we look at the tropical regions only, Van Minnen et al. (2009) suggest that prior to 1950 these regions were neutral in terms of carbon release, but the rapid increase in deforestation after that resulted in these regions being major carbon sources. Although, energy and industry related emissions are currently increasing in many tropical nations, they still account for less than the land use emissions (Van Minnen et al. 2009).

Estimating the carbon stock changes in these forests is an extremely challenging task, because it requires precise land cover maps, accurate biomass estimates as well as knowledge on the fate of carbon (Prentice et al. 2001). Global land cover maps are derived from the interpretation of several satellite images and commitments need to be made in order to get a product within a reasonable time frame. This means that high-resolution satellites can be used to get more precise maps; however, the time (and money) it would take to interpret several land cover transitions in thousands of images for the whole globe makes it unfeasible. Instead lower resolution satellites are used, but this has the cost of losing the possibility of measuring land cover changes precisely. Moreover, although satellite image is an amazing tool for Earth observation, the outputs still have strong uncertainties attached as the interpretation of satellite images faces difficulties such as separating similar land cover types (*e.g.* forest *versus* regeneration, deforestation *versus* degradation) as well as persistent cloud cover. Very high resolution sensors developed in the last few years (Asner et al. 2013) might help overcome some of these problems, but historical databases with low uncertainties attached of global land cover changes still seem far from reality.

Maps of carbon content of plants in the terrestrial ecosystems, and in particular in tropical rainforests, can be created either using a bottom-up (forest inventories or ecological studies) or a top-down (satellite imagery interpretation) approach. Usually, in forest inventories and ecological studies (Labrecque et al. 2006) carbon stocks of vegetation (and soils) is measured and then aggregated to a larger scale. However, high spatial and temporal heterogeneity and methodological differences within years introduce large uncertainties. As described above, products derived from satellite image interpretation (Baccini et al. 2008) also present high uncertainties and have their own inherent difficulties. The best option is to combine both methodologies (Saatchi et al. 2011; Baccini et al. 2012) taking advantage of each strengths and reducing their weaknesses.

Finally, to know exactly the fate of carbon after land conversion, including which fraction of biomass is burned, the rates of decomposition of the different biomass components, as well as the impact of burning on soil carbon, and subsequent land cover, is an extremely challenging task due to the spatial heterogeneity of the tropical deforestation process (Prentice et al. 2001). However, delayed carbon fluxes, from biomass decomposition for example, implies that estimates of current fluxes must include data on historical land cover activities and associated information on the fate of cleared carbon (Houghton et al. 2012).

Several attempts have been made to determine the amount of carbon released to the atmosphere due to land use change. Such estimates rely heavily on the land cover data sets and estimates of average carbon density as described above, which means they are not as accurate as estimates of fossil fuel emissions. About 10 to 30% of the current total anthropogenic emissions of CO_2 are estimated to be caused by land-use conversion (Prentice et al. 2001). The expansion of croplands at the expense of natural ecosystems as well as the conversion of forest into cattle pasture continues to dominate the carbon emissions from the tropical biosphere (Malhi et al. 2002). Using a simple model based on deforestation rates statistics and several pools and rates of regrowth, (Houghton et al. 1999; Houghton et al. 2000) estimated carbon emissions to be 121 Pg C between 1850 and 1990. From this approximately 60% were due to land cover changes in tropical areas, whereas the remaining 40% were due to changes in temperate areas, reflecting the importance of tropical deforestation in the global land use change carbon budget. Furthermore, the same author estimated an annual flux of carbon from land-use change for the period from 1990 to 1995 to be 1.6 Pg C/yr, consisting of a source of 1.7 Pg C/yr in the tropics and a small sink in temperate and boreal areas (Houghton et al. 2000). These results are consistent with those found by Malhi et al. (2002), who determined that, tropical forests of Southeast Asia released 0.76 Pg C/yr in 1990s, whereas Africa rainforests released 0.34 Pg C/yr and South American tropical forests emitted 0.50 Pg C/yr in the same period.

In the last few years forest protection measures that rely on carbon emissions have been discussed and to some extent implemented (*e.g.* REDD by the United Nations collaborative initiative on Reducing Emissions from Deforestation and forest Degradation). These type of measures aim to financially compensate tropical nations, that avoid carbon emissions due to deforestation in their territory (Corbera et al. 2010). As shown above, researchers are still faced with many difficulties to accurately estimate carbon losses due to tropical deforestation. Reducing uncertainties attached to carbon emissions estimates from land cover change is vital for the success of conservation policies which can help to secure the future of tropical forests.

5) Thesis outline

The main objective of my PhD was to develop a spatially-explicit, predictive and dynamic model of tropical deforestation. This model can then be used to test different scenarios which can help predicting future impacts of land cover change and understand how implemented policies can affect the patterns and rate of tropical deforestation. The model is strictly data dependent and adaptable for use at different scales and in different regions. After the Introduction (Chapter 1) I present four chapters that compile the research done throughout my PhD. All chapters are presented in the form of manuscripts already (or in the process of being) submitted to peer-reviewed international journals. The outline of this thesis is as follows:

- Chapter 1 – Introduction (above)

- Chapter 2 – Changes in size of deforested patches in the Brazilian Amazon

This chapter of my PhD thesis includes the manuscript entitled "Changes in size of deforested patches in the Brazilian Amazon" published in *Conservation Biology* in 2012. In this study the proportion of deforestation associated with different-sized clearings in the Brazilian Amazon was analysed from 2002 through 2009. Annual deforestation maps were used to investigate how deforestation patch sizes varied in recent years, by determining total area deforested in a given year and the size distribution of patches deforested in that year.

- Chapter 3 – The transparency, reliability and utility of land-use and land-cover change models: an Amazonian case study

This chapter of my PhD thesis includes the review manuscript entitled "The transparency, reliability and utility of land-use and land-cover change models: an Amazonian case study" currently under review in *Global Change Biology*. The aim of this review is to give a summary and compare modelling methodologies (*e.g.* study resolution, calibration and validation methods used) used to model land cover change in the Amazon, and to quantitatively compare model outputs as well as, whenever possible, validate them against real data. This manuscript was written in collaboration with Sadia E. Ahmed.

- Chapter 4 – Predictive modelling of contagious deforestation in the Brazilian Amazon

The fourth chapter of my PhD thesis includes the manuscript entitled "Predictive modelling of contagious deforestation in the Brazilian Amazon" which has been accepted for publication in *PLOS One* in September 2013. A spatially-explicit predictive model of deforestation was developed for the Brazilian Amazon, which differs from previous models in three ways: (1) it is probabilistic and quantifies uncertainty around predictions and parameters; (2) the overall deforestation rate emerges 'bottom up', as the sum of local-scale deforestation driven by local processes; (3) deforestation is contagious, such that local deforestation rate increases through time if adjacent locations are deforested.

- Chapter 5 – Evaluating the accuracy of land cover change models: sensitivity to calibration year and temporal changes to deforestation processes

In the fifth chapter of this thesis the manuscript entitled "Evaluating the accuracy of land cover change models: sensitivity to calibration year and temporal changes to deforestation processes", currently under review in *Regional Environmental Change* is presented. Here, an analysis using a large historical land cover change (LCC) database, encompassing nearly three decades in the municipality of Machadinho d'Oeste in the Brazilian Amazon, was performed to: (1) investigate how parameter values associated with drivers of LCC vary through time; (2) determine whether the parameter values associated with different land cover transitions (forest to deforested, regeneration to deforested or deforested to regeneration) are more or less variable than each other; and (3) quantify the influence of choosing a particular time period for model calibration on the accuracy of LCC model predictions.

- Chapter 6 – The carbon legacy of modern tropical deforestation

The final data chapter of my PhD thesis present an analysis where the probabilistic model of deforestation developed in Chapter 4 was combined with a terrestrial carbon model, and a widely used bookkeeping carbon model to estimate carbon emissions due to gross deforestation from 1950 to 2009 for the Amazon, the Congo Basin and Southeast Asia. This carbon legacy corresponds to the amount of carbon yet to be released to the atmosphere due to past deforestation. This manuscript it being prepared to be submitted to *Nature*.

- Chapter 7 – Discussion

In the closing section of this thesis a thorough global discussion of my PhD project is presented, including its main contributions to this research field (land-cover/land-use change modelling) and some insights on where I believe efforts should be made to improve this research field in the future.

Chapter 2 – Changes in size of deforested patches in the Brazilian Amazon

Isabel M. D. Rosa[1], Carlos Souza Jr[2]. and Robert M. Ewers[1]

[1] Imperial College of London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, United Kingdom

[2] Imazon – Amazon Institute of People and the Environment, Rua Domingos Marreiros, 2020, Fátima, CEP: 66.060-160, Belém, Pará, Brazil

Manuscript submitted to *Conservation Biology* in July 2011

Revised manuscript accepted in April 2012: *Conservation Biology*, Volume 26, No. 5, 932–937

Keywords: conservation effectiveness, size distribution, spatial dynamics, temporal dynamics, tropical deforestation

Running title: Dynamics of Amazonian Deforestation

Corresponding author: Isabel MD Rosa (i.rosa09@imperial.ac.uk)

2.1) Abstract

Different deforestation agents, such as small farmers and large agricultural businesses, create different spatial patterns of deforestation. We analyzed the proportion of deforestation associated with different-sized clearings in the Brazilian Amazon from 2002 through 2009. We used annual deforestation maps to investigate how deforestation patch sizes varied in recent years, by determining total area deforested in a given year and the size distribution of patches deforested in that year.

The size distribution of deforested areas changed over time in a consistent, directional manner. Large clearings (>1000 ha) comprised progressively smaller amounts of total annual deforestation. The number of smaller clearings (6.25-50.00 ha) remained unchanged over time. Small clearings accounted for 73% of all deforestation in 2009, up from 30% in 2002, whereas the proportion of deforestation attributable to large clearings decreased from 13% to 3% between 2002 and 2009. Large clearings were concentrated in Mato Grosso, but also occurred in eastern Pará and in Rondônia. In 2002 large clearings accounted for 17%, 15%, and 10% of all deforestation in Mato Grosso, Pará, and Rondônia, respectively. Even in these states, where there is a highly developed agricultural business dominated by soybean production and cattle ranching, the proportional contribution of large clearings to total deforestation declined. By 2009 large clearings accounted for 2.5%, 3.5%, and 1% of all deforestation in Mato Grosso, Pará, and Rondônia, respectively.

These changes in deforestation patch size are coincident with the implementation of new conservation policies by the Brazilian government, which suggests that these policies are not effectively reducing the number of small clearings in primary forest, whether these are caused by large landholders or smallholders, but have been more effective at reducing the frequency of larger clearings.

2.2) Introduction

The Brazilian Amazon is the largest tropical rainforest in the world, with an area of about 4 million km^2 (Pereira et al. 2010a), and it has the highest area of forest cleared per year (Nepstad et al. 2009). Annual deforestation averages 0.5%/year (FAO 2010). Estimates of the mean rate of annual deforestation vary from 16,300 km^2/yr between 2001 and 2007 (Loarie et al. 2009) to 19,500 km^2/yr between 1996 and 2005 (Nepstad et al. 2009). Recently deforestation rates in the Brazilian Amazon decreased to an average 13,000 km^2/yr between 2005 and 2009 (INPE 2011).

Deforestation rates in the Brazilian Amazon exhibited high annual variability over the past 20 years. These fluctuations were related to the economic status of the country, infrastructure development, and global demand for agricultural products such as beef or, more recently, soybeans (Fearnside 2005; Morton et al. 2006; Nepstad et al. 2006a; Ewers et al. 2008). Before the 1960s, deforestation of the Brazilian Amazon was limited mainly by its isolation. As a result of infrastructure development and government support of agriculture and colonization schemes, this pattern changed in the 1970s and 1980s and there was a rapid increase in the area deforested (Rudel 2005; Nepstad et al. 2006a).

The government has since reduced or eliminated direct sponsorship of deforestation. Deforestation is now mainly driven by the growing demand for timber and agricultural products, the profitability of which has been influenced by macroeconomic shocks in the Brazilian economy and changes in international exchange rates (Ewers et al. 2008; Rudel et al. 2009). Between 1997 and 2004, the expansion of the agricultural industry that accompanied the growing Brazilian economy led to a continuous increase in deforestation rates (Nepstad et al. 2006a). However, deforestation rates have fallen by over 25% since 2004. This fall is believed to be related to the global economic recession, which led to a drop in soybean prices and the implementation of conservation policies and law enforcement by the Brazilian government (Nepstad et al. 2009).

Specific conservation actions in the last decade have included the expansion of protected areas, command-and-control operations, and listing of critical municipalities (those with high rates of deforestation) in which access to rural credit by landowners is restricted. In

2004 the Brazilian government implemented the Plan to Prevent and Combat Deforestation in the Amazon (Plano de Prevenção e Combate ao Desmatamento da Amazonia, PPCDAM). Since 2005 Brazilian Institute of Environment and Renewable Natural Resources (IBAMA) has also implemented a series of command and control operations (Operação Arco de Fogo, Operação Guardiões Termópilas, and Operação Disparada) to combat illegal deforestation in which they obtain early alerts about the location of illegal deforestation from the Real Time Deforestation Detection System, developed by INPE. Market campaigns led by Greenpeace established a soy moratorium in 2006, and, more recently, a cattle moratorium, led by the Federal Prosecutor Office in Pará, was established. Both moratoriums committed to the decommercialization of these products originating from newly deforested areas in the Brazilian Amazon.

Deforestation rates can be extremely sensitive to the national political situation. For example, in early 2011 a rumour (that was shortly after confirmed) started that the Brazilian Chamber of Deputies would vote to ease rules on conserving Amazon rainforest and would forgive past illegal deforestation. By April a sharp increase in the amount of land cleared was recorded; deforestation rose by over 500% relative to the same period the year before (Hayashi et al. 2011).

Our objectives were to analyze how much deforestation was associated with clearings of different sizes in the Brazilian Amazon and to determine whether the distribution of clearing sizes has changed over time. Quantifying temporal changes in the size distribution of deforested patches may help determine whether the conservation policies of the Brazilian government are reducing deforestation rates.

2.3) Methods

We used INPE's annual deforestation maps for 2002 through 2009 (Camara et al. 2006) to investigate how deforestation patch sizes varied in recent years. Although 2001 data were available, we did not use them because 2001 was the first year in which INPE was able to obtain full satellite coverage of the Brazilian Amazon (INPE 2012a). Therefore, deforestation events detected in 2001 reflect deforestation that occurred in that year and in previous years. Landsat images taken during the dry season that cover the Brazilian Amazon are used to compile the deforestation maps. Deforestation is defined by INPE as the anthropogenic conversion of primary forest for agricultural and cattle production. They use orbital platforms to detect such deforestation (Camara et al. 2006); therefore, clearing of secondary forests is not detected. The main limitation of these data is that the spatial resolution of the images does not allow the identification of deforested patches smaller than 6.25 ha. This means we were unable to include in our analyzes the proportion of area cleared by many small landholders who practice slash and burn farming and deforest on average 2-3 ha/year (Fujisaka et al. 1996; Metzger 2002).

We used the annual deforestation maps to determine total area deforested in a given year and the size distribution of patches deforested in that year. We classified all patches deforested in a given year into 1 of 6 size classes (6.25-50 ha, >50-100 ha, >100-200 ha, >200-500 ha, >500-1000 ha, and >1,000 ha) (Alves 2002). We defined a deforested patch as a contiguous area of forest that was cleared within 1 year and used linear regression to identify trends in size of deforested patches over time. We also aggregated the annual data to map the spatial distribution of sizes of deforested patches in a period of rapidly increasing deforestation (2002 through 2004) and in a period of rapidly decreasing deforestation (2005 through 2009).

A shift in the size distribution of deforested patches could mean past deforestation was of sufficient extent to limit new deforestation to small extents. To investigate this possibility, we used INPE's forest cover and deforestation data from 2002 through 2009 in Pearson's correlations (r) to determine the correlation between the median sizes of deforested and forest patches within each of the 412 municipalities for which we had 8 years of data.

2.4) Results

From 2002 through 2009 there was a continuous increase in the proportion of deforestation associated with small (6.25-50ha) clearings (β= 6.12, t=15.93, R^2=0.98, df=6, p<0.001): from 30% in 2002 to 73% in 2009 (Fig. 4a). Simultaneously, there was a continuous reduction in the proportion of deforestation associated with large (>1,000 ha) clearings (β=-

1.77, t=-6.84, R^2=0.89, df=6, p<0.001), from 13% in 2002 to 3% in 2009 (Fig. 4a). From 2002 to 2009 there were 90% fewer new clearings of ≥50 ha. By contrast, the total area of deforestation associated with small clearings decreased 35% over the same period, meaning that the decrease in deforestation rate from 2006 through 2009 was the result of a sharp reduction in the area deforested patches of ≥50 ha (Fig. 4b).

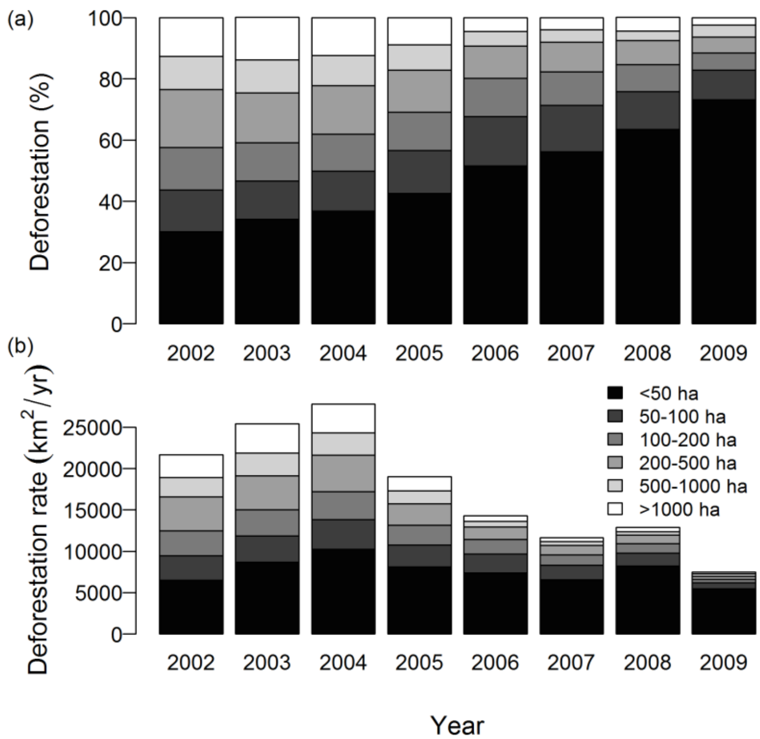


Figure 4 – (a) Percentage and (b) area of deforested patches of different sizes in the Brazilian Amazon from 2002 through 2009.

The spatial distribution of sizes of deforested patches did not change when deforestation was increasing (2002-2004) or decreasing (2005-2009). In both periods, Mato Grosso, Pará, Rondônia, and Amazonas had a higher proportional area cleared in large patches than Acre, Amapá, Maranhão, Roraima, and Tocantins (Fig. 5). Even in states with highly developed agricultural businesses dominated by soybean production and cattle ranching, such as Mato Grosso, Pará, and Rondônia, the proportion of deforestation attributable to large clearings declined from 17%, 15%, and 10% in 2002 to 2.5%, 3.5%, and 1%, respectively, in 2009.

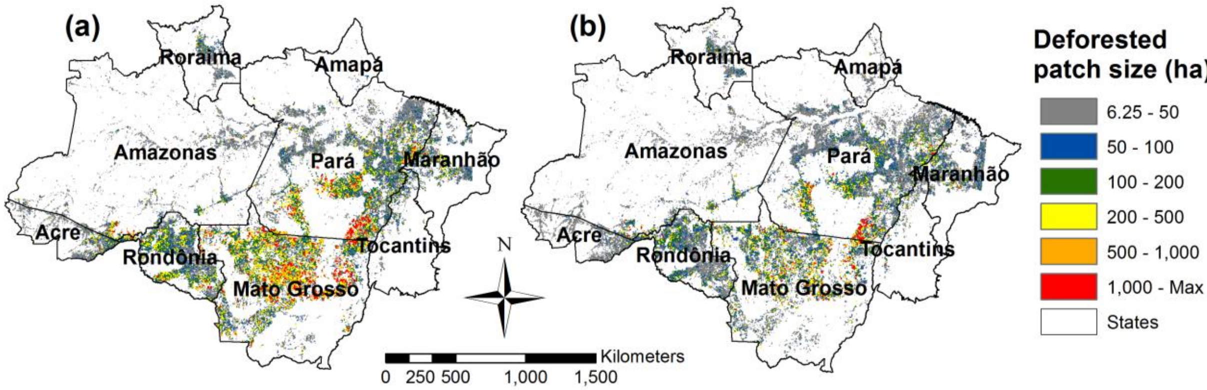


Figure 5 – Distribution of deforested patches of different sizes in the Brazilian Amazon for periods of (a) rapidly increasing deforestation (2002 through 2004) and (b) rapidly decreasing deforestation (2005 through 2009).

For most municipalities (88%), the sizes of deforested and forested patches were not significantly correlated (Fig. 6). Significant values ($r \geq 0.707$, df=6, $p \leq 0.05$) suggest the size of deforested patches may be limited by the amount of remaining forest, whereas non-significant values ($r < 0.707$, df=6, $p > 0.05$) suggest there is no apparent limit to the potential size of deforested patches. The small number (12%) of municipalities with significant positive correlations (mean $r = 0.80 \pm 0.02$ 95% C.I., df=6, $p < 0.05$) were along the Arc of Deforestation, where extensive deforestation is related to proximity to urban centres and roads and a more suitable climate for agriculture (Aguiar et al. 2007).

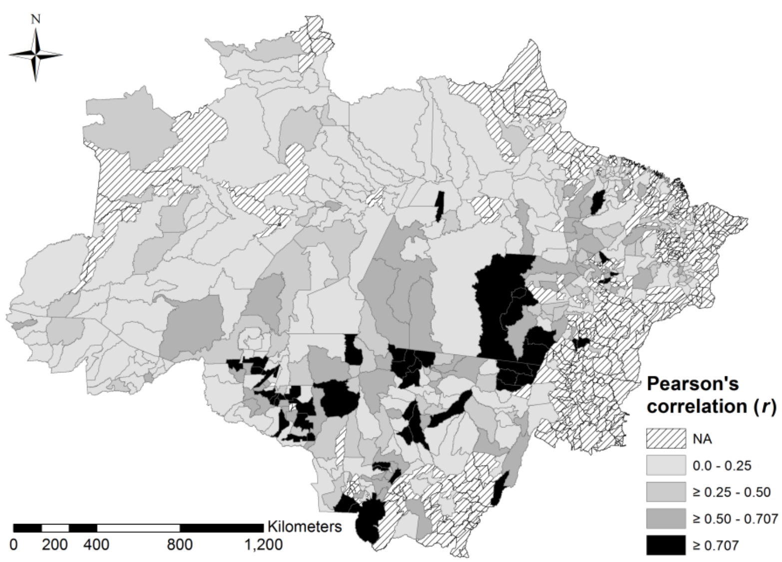


Figure 6 – Pearson's correlation coefficient (r) between median sizes of deforested and forest patches in the Brazilian Amazon from 2002 through 2009.

2.5) Discussion

The size distribution of deforested patches in the Brazilian Amazon is changing in a consistent, directional manner. Large clearings (>1,000 ha) comprised progressively smaller amounts of total annual deforestation, whereas the proportion of smaller clearings (6.25-50.00 ha) remained unchanged over time. Even in states where agricultural businesses are highly developed and dominated by soybean production and cattle ranching, the proportional contribution of large clearings to total deforestation declined from 2002 through 2009.

The number of larger clearings (Fig. 4b) decreased coincident with the implementation of Plan for Protection and Control of Deforestation in the Amazon (PPCDAM). Therefore, we believe PPCDAM, the global economic crisis, and market campaigns (e.g., the Soy Moratorium) aimed at decreasing deforestation - affected patterns of deforestation. Clearings of >100 ha are a priority (for PPCDAM, and our results show a reduction in the number of clearings in this size class (Fig. 5b). Our results, then, provide support for the suggestion made by Michalski et al. (2010) that smallholders are gradually accounting for a greater proportion of annual deforestation in the Amazon. However, recent changes in the size distribution of deforested areas were coincident with a global economic downturn, which has also likely contributed to the reduced rate of clearing of large areas.

Moreover, we found evidence that in some municipalities historic deforestation is limiting the extent of present-day deforestation. The size of the deforested patches may be decreasing because landholders are changing their behaviour. Owners of large tracts of land may be avoiding creating large clearings to prevent a response from the local or federal government and instead may be clearing many small patches (Michalski et al. 2010). In the absence of Amazon-wide data on property boundaries, this hypothesis cannot be systematically tested. The INPE data we used documents the conversion of primary forests only, meaning that large clearings of secondary forests may not be detected.

The reduction in the number of large clearings from 2002 through 2009 means that small clearings now account for 80% of annual deforestation in the Brazilian Amazon. This result contrasts with results of studies conducted before the widespread use of remote sensing for analyses of deforestation and suggests that spatially extensive cattle ranching operations account for >50% of all deforestation (Fearnside 1993; Walker et al. 2000; Chomitz & Thomas 2001). More recently, Ferraz et al. (2005) suggested that deforestation is concentrated in small patches (6.25-50.00 ha), and Pacheco (2009) found that smallholders accounted for 35% of all deforestation in 2003. However, we found that averaged across all years from 2002 through 2009, small patches accounted for almost 43% of all deforestation whereas large patches (>1000 ha) accounted for 10% of deforestation. The proportion of deforestation attributable to small clearings now is much greater than that of large clearings and is likely to be higher than we report here because the INPE data on which we based our analyses does not detect deforested patches <6.25 ha.

Land-tenure legislation affects the property rights of many farmers in the Brazilian Amazon with small landholdings (Barreto et al. 2008). Insecure property rights can lead property owners to manage their land for short-term profit rather than long-term sustainable income (Rudel 2005). The incentive to invest in land improvement and land maintenance is low if land owners have no guarantee of owning that land in the future (Alston et al. 1996; Svensson 1998; Araujo et al. 2009).

With a decrease in the average deforestation rate from 2005 through 2009 of about 67% relative to deforestation form 1996 through 2005, Brazil is not far from achieving an 80% reduction in deforestation rate by 2020. Because PPCDAM focuses on large clearings, it is unlikely that PPCDAM alone will be able to reduce deforestation rates below their current level. Our results show that clearings below the 100-ha threshold on which PPCDAM focuses accounted for 0.5 million ha of deforestation in 2009. Consequently, even if no clearings are >100 ha, deforestation rates will still be 1,000 km^2/yr higher than the targeted 80% reduction by 2020 (Fig. 5b). We believe that developing a stronger property-rights system while improving the spatial resolution of deforestation surveillance techniques will further decrease rates of deforestation in the Brazilian Amazon to the lowest levels in 30 years.

2.6) Acknowledgments

This work was funded by the Grantham Institute for Climate Change and Microsoft Research Cambridge.

Chapter 3 – The transparency, reliability and utility of land-use and land-cover change models: an Amazonian case study

Isabel M.D. Rosa[1*]**,** Sadia E. Ahmed[1*] and Robert M. Ewers[1]

[1]Imperial College of London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, United Kingdom

* Equal contribution/joint first authors

Manuscript submitted to *Global Change Biology* in December 2012.

Revised manuscript submitted to *Global Change Biology* in April 2013.

Keywords: Amazon, input data, LULC change, model comparison, modelling framework, spatio-temporal scales, validation methods

Running title: Modelling land-use and land-cover change

Corresponding author: Isabel MD Rosa (i.rosa09@imperial.ac.uk)

3.1) Abstract

Land use and land cover (LULC) change is one of the largest drivers of biodiversity loss and carbon emissions globally. Tropical deforestation is the most paradigmatic example of LULC change, as such we use the tropical Amazon as a case study to investigate predictive models of LULC change. Current predictions differ in their modelling approaches, and predictions of future change are highly variable and often poorly validated. We carried out a quantitative review of 35 modelling methodologies, considering model spatio-temporal scales, inputs, calibration and validation methods. In addition, we requested model output data from each of the models reviewed and carried out a quantitative assessment of model performance for LULC predictions in the Brazilian Amazon.

We highlight existing shortfalls in the discipline and uncover three key points that need addressing to improve the transparency, reliability and utility of LULC change models: 1) a lack of openness with regard to presenting and making available the model inputs, model code and model outputs, 2) the difficulties of conducting appropriate model validations, combined with a stronger recognition of the importance of validating model outputs, and 3) no standardised framework that can be used as a basis for comparing LULC model predictions and generating multi-model inference.

We further draw comparisons between LULC change models and climate change modelling attitudes and paradigms. We suggest that the rise of climate change models provides a pathway that LULC change modellers may emulate to greatly improve the discipline. Climate change models have exerted considerable influence over public perceptions of climate change and now impact policy decisions at all political levels. We suggest that LULC change models have an equally high potential to influence public opinion and impact the development of land use policies based on plausible future scenarios, but to do that requires a step-change in the discipline.

3.2) Introduction

Land use and land cover (LULC) change is a process that is present in all environments across the globe (Lambin et al. 2001; Geist & Lambin 2002). It is driven by many natural and anthropogenic factors and is the largest driver of biodiversity loss at global scales (Pereira et al. 2010b). There are many models that attempt to predict LULC changes at all spatial scales. Some models work at global scales (Nelson et al. 2010; Pereira et al. 2010b) and use large scale drivers that are applicable to all systems, such as distance to infrastructure and population growth. Other models work at regional scales such as models of Central or South America (Wassenaar et al. 2007a), at country scale (Lapola et al. 2010), or even down to local scales such as modelling a single reserve like the Xingu National Park in Brazil (Maeda et al. 2011). Overall, it is tropical LULC change that has received the most attention, and in particular models of tropical deforestation are prevalent in the literature. Tropical deforestation is probably the most paradigmatic example of LULC change, because of the huge detrimental impacts forest loss can have on the future of the planet and human wellbeing (Foley et al. 2005). During the last two decades, 80% of new agricultural land across the world has been a result of tropical forest destruction (Gibbs et al. 2010). Furthermore, emissions from global land use change are the second-largest anthropogenic source of carbon dioxide (CO_2), just behind fossil fuel emissions, with Southeast Asia and South America being the two main contributors (Le Quere et al. 2009). Losing biodiversity-rich ecosystems at such a fast rate is a major threat to the world's biodiversity (Myers 1988). In addition, the rapid destruction of tropical forests is compromising the future of many indigenous people (Alcorn 1993) as well as the future of local populations who use forests as a source of food, construction materials, remedies, and who also sell forests products at local markets for income (Laurance 1999).

Numerous models of tropical LULC conversion have been developed to understand the complex interactions among human and biophysical factors that drive change (Ludeke et al. 1990; Mertens & Lambin 1997; Verburg et al. 2002 ; Soares-Filho et al. 2006). LULC models are employed to address questions concerning why changes happened in the past, to help understand the main drivers of change in the present, to predict how much and where change will occur in the future, and to examine plausible scenarios of landscape modification.

Predicting not only the amount of forest that will be lost in the future, but also the location of this loss, is vital to successfully implementing conservation strategies (Mertens & Lambin 1997). Current predictions of LULC change differ in their modelling approaches, and predictions of future change are highly variable and are often poorly validated. Thus, modelling LULC change processes remains a great challenge. This challenge arises partly because the physical environment can vary greatly from one region to another, and can also be in constant change. In addition, the underlying processes that drive LULC change are usually very complex, combining many socio-economic, cultural, political and environmental factors (Geist & Lambin 2002). In the literature there are a variety of predictive LULC change models, which vary greatly in terms of methodology (*e.g.* agent-based, cellular automata, statistical), time frame, and the region where, and scale at which, they were calibrated There are also, however, common features among the models, such as a universal reliance on data that is mainly derived from satellite image interpretation or other geographic information systems.

The Amazon has been the focus of many LULC modelling endeavours because it is the largest continuous area of remaining tropical forest, which provides many ecosystem services, both locally and globally (Moran 1993; Foley et al. 2007; Betts et al. 2008; Bonan 2008). The Amazon covers approximately 6 million km^2, crossing nine nations' boundaries, of which Brazil hosts the largest portion with approximately 60% of the total area. However, 15% of the forest cover in the Brazilian Amazon had been deforested by 2009 (Pereira et al. 2010a). Deforestation in South America rapidly increased in the late 1960s and early 1970s when governments started to end the passive protection of the Amazon, when previously inaccessible areas became accessible with the development of road networks into frontier areas (Fearnside 2005; Armenteras et al. 2006). Colonisation schemes were implemented which featured extensive road building, and land settlement schemes along the newly constructed highways (Steininger et al. 2001 ; Fearnside 2005; Killeen et al. 2007). The agro-industrial expansion, especially for soybean production and cattle ranching, associated with a healthy and growing economy in the region, led to a continuous increase in deforestation rates through to the mid-2000s (Nepstad et al. 2006b). More recently, deforestation rates have fallen, led primarily by reduced rates of forest loss in Brazil. This fall is believed to be related to the global economic recession (Nepstad et al. 2009), which led to a drop in the soy prices, as well as with successful conservation (expansion of protected areas network and market based strategies) and law enforcement (control and command field operations) policies applied by the Brazilian government (Rosa et al. 2012).

This review provides a summary of predictive models of LULC change, using the Amazon as a case study. We aim to highlight the different methodologies that exist, specifically with regards to differences in prediction goals, model inputs and outputs, and model calibration and validation techniques. We achieve this goal through a quantitative review of LULC change models operating in the Amazon region. Based on the findings of this analysis, we highlight several shortcomings in the approaches taken to LULC change modelling, and draw on the experience of the climate change modelling fraternity to make specific recommendations with a view to strengthening the reliability of the LULC change modelling discipline.

3.3) Material and Methods

Using ISI Web of Knowledge, we searched for papers using the keywords 'Amazon land use change model', on the 9th of September 2012, which returned a total of 548 papers. From this we selected a set of primary papers (35 in total) that specifically model LULC change in the Amazon (Appendix A1). To test for the transparency of this set of LULC change models, we extracted the methodological information from these 35 papers to conduct a quantitative review of model approaches, covering aspects such as (1) the spatial (cell size and extent of study area) and temporal (time period for which the model was calibrated and simulation years) scale of models (Table A1 in Appendix A). We assessed the correlation between the model extent and the cell size used, and calculated basic statistics to identify trends in the time period of models; (2) model type (e.g. cellular automata, agent-based); and (3) data inputs used. Models that lack transparency are those for which we were unable to extract the information described above. We assessed the reliability of models by (4) examining the methods used to calibrate and validate the models, and finally, the utility of models was determined by (5) our ability to obtain the modelled predictions in a form that could be used by other researchers and decision makers.

We classified models into one of five categories: (1) models that were based on the decisions of LULC change agents were considered “Agent-based” (Parker et al. 2008); (2) models that accounted for the neighbourhood when determining change were defined as “Cellular automata” (White & Engelen 2000); (3) models purely based on the extrapolation of past trends were defined as “Statistical” (Millington et al. 2007); (4) models developed with the goal of optimising income or minimising losses were considered “Optimisation” (Chuvieco 1993); and (5), models that used other algorithms to identify trends were defined as “Other” (Table A1 in Appendix A). We used a Chi-squared test to identify any significant bias in the type of models used. Finally, we categorised models as being deterministic or stochastic. Deterministic models use inputs and create outputs that are fixed, meaning the same model run multiple times will always give the same result. By contrast, stochastic models use inputs that are described by a probability distribution of some description and so contain a degree of randomness that can be used to estimate the level of uncertainty around model predictions.

We define model inputs as the factors or parameters that a model takes into account to make predictions. Landscape change modelling often uses many inputs because models are attempting to replicate the inherently complicated phenomena of future LULC change, which is heavily influenced by human behaviour. As such, we divided model inputs into four broad categories; (1) geographical, (2) economic, (3) social, and (4) biological inputs (Table A1 in Appendix A). Geographical inputs play a vital role in LULC change modelling, providing the environmental setting that describes the real world on top of which the model can make predictions. Economic inputs cover factors relating to monetary gains and losses, for example the amount of capital available or land prices. Social inputs consider what people value, how people live, and include factors such as family size and family demography. Biological inputs are used to predict the utility of converting land from forest to another land use, using soil fertility for instance (Carpentier et al. 2000). Model inputs were also divided into categories according to whether they are static or dynamic inputs. Static inputs differ from dynamic inputs in that they do not change through time in the model. For example, the location of key cities or topographical patterns can be considered static over the time periods modelled. By contrast, dynamic inputs are continuously updated within the model itself (Appendix A).

To assess the reliability of the 35 models, we recorded how the model calibration and validation were carried out, as well as how model outputs were validated against observed data. Calibration is formally defined as "the estimation and adjustment of model parameters and constants to improve the agreement between model output and a data set" (Rykiel Jr. 1996). The process of validating and assessing a model's predictive power involves comparing the model predictions against observed data (Table A1 in Appendix A). We attempted to conduct a series of standardised validation tests on published LULC change models, so we requested via e-mail digital maps of model predictions from the authors of each model we considered in this review. We e-mailed the corresponding author of each paper up to three times, and if we received no response we e-mailed the co-authors for which we were able to find e-mail addresses. Models for which we were unable to obtain the predictions represent models that have only limited utility for decision makers, who will typically require access to detailed spatial information about projected LULC changes.

For the models that we were able to obtain model predictions in a format that could be compared with reliable observed data, we made quantitative comparisons of the model outputs and accuracy. These comparisons were made initially on a pixel-by-pixel basis using three simple measures: (1) 'match', representing the proportion of deforested area correctly predicted by the model when compared to observed deforestation; (2) 'omission', representing false negatives (model predicted no deforestation in a location where deforestation occurred); and (3) 'commission', representing false positives (deforestation was predicted but did not occur). Using annual deforestation maps for the Brazilian Amazon (INPE 2012b) we created binary raster files representing annual (deforestation in that year) and cumulative (accumulated deforestation that occurred between 2002 and that year) observed deforestation from 2002 through 2010. For each model that we validated, we constructed different raster files to match the spatial extent and resolution of the observed data to that used in model predictions. Then, using the raster maps of deforestation predictions collected from the authors, we compared on a pixel-by-pixel basis where deforestation was perfectly predicted (= match), omitted or committed. This was done for all pixels in the landscape and we summed all the pixels that matched and divided the sum by the amount of change observed to get the percentage match. There is a general reluctance to use pixel-by-pixel comparison methods for LULC model validation because there is no differentiation between 'near miss' and 'far miss' errors (Pontius et al. 2002, Pontius et al. 2004). We agree that these simple metrics (match, omission and commission) represent extremely stringent tests of model reliability, but we also argue that they represent exactly the ability of LULC models to predict the spatial patterns of LULC change, and ably represent the two cases in which those predictions can be wrong. To allow for near and far misses, we also calculated a distance-based measure of model match, annually and cumulatively, by defining a set of buffer zones (1, 5, 10, 50 pixels in radius) around each pixel of predicted deforestation, and calculated the proportion of observed deforestation that was found within those buffers. We used pixels rather than distance, as pixel size was correlated with model extent and therefore represents a standardised metric of scale that accounts for the differences in model extent among the three models we compared. This distance-based validation metric quantifies the degree of spatial error in model predictions.

3.4) Results

3.4.1) Spatial and temporal scales

Of the 35 models, only three covered the whole Amazon basin; 19 were applied only in the Brazilian Amazon, six were in Ecuador and the remaining seven models were distributed among Colombia (2), Peru (2) and Bolivia (3) (Fig. 7a). Within the Brazilian Amazon, most sub-regional models (those that do not cover the whole Brazilian Amazon) were developed for states situated in the so-called "Arc of Deforestation" (Fig. 7b), particularly Mato Grosso (5), Pará (6) and Rondônia (3). This is the region of the Brazilian Amazon where there is a very active deforestation frontier, due to the easy access to forests and the reduced transportation costs provided by the well-developed road network (Laurance et al. 2002; Fearnside 2005; Aguiar et al. 2007).

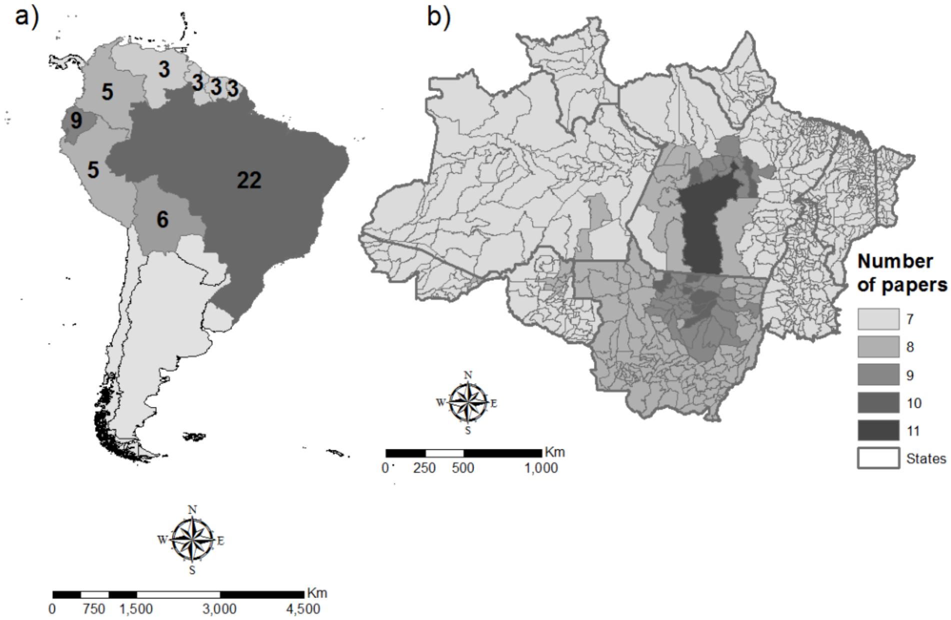


Figure 7 – Number of papers included in our review (a) per country in South America and (b) within the Brazilian Amazon.

The spatial extent of LULC change models is generally biased toward regional-scale models (median = 23,500 km^2, ranging from about 300 km^2 to more than 8 million km^2), and the spatial scale of models was closely correlated with the resolution, or cell size, of models (Pearson's correlation on log extent (km^2) with log cell size (km^2); $r = 0.78$, df = 22, $p < 0.001$) (Fig. 8). Most Amazonian LULC change models select one particular scale at which to work (Supporting Information), and we identified just one paper that operated at multiple scales, integrating small-scale and regional-scale modelling approaches through a combination of linked models (Moreira et al. 2009).

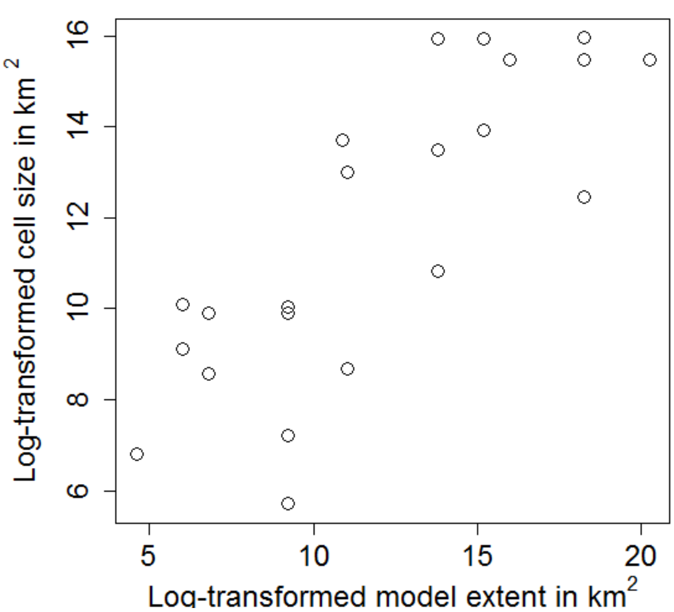


Figure 8 – Correlation between the spatial scale of models (model extent in km^2) and the model resolution (given by the cell/pixel area in km^2).

We detected large amounts of variation in the temporal scale over which models were used to predict future LULC changes. Of the 35 models, three did not provide any future predictions apart from the initial year, while on average the other 32 models made predictions extending 24 years into the future, ranging from just 5 years (Mello & Hildebrand 2012) to a maximum of 50 years (Evans et al. 2001). The distribution of number of years of change was left-skewed, meaning that most papers tend to focus on short and medium temporal extents. This is likely due to a perceived tendency of model predictions to become increasingly uncertain into the future due to the very large number of dynamically adjusting variables that cannot be accurately accounted for in models (Deadman et al. 2004).

3.4.2) Model Type

Across the set of 35 papers we identified five broad categories of model types: agent-based (n=12), cellular automata (n=10), statistical (n=13), optimisation (n=3), and other types of models (n=6). Overall, there is no significant bias towards any particular type of model (Chi-squared test, $\chi^2 = 8.05$, df = 4, p = 0.09). Some models fell into more than one category, such as optimisation models that were often combined with agent-based models, allowing models to either maximise gains or minimise losses from the farmers' (agents') perspective. We found that agent-based models were most commonly used when modelling LULC change at local and regional scales (mean extent = 13,000 km^2 ± 10,000, 95% C.I.), whereas cellular automata were more commonly used for large scale models (mean extent = 2,770,000 km^2 ± 2,460,000, 95% C.I.). Of the six 'other' types of model, the methods implemented included the use of Markov chains and neural networks to train models to gain an understanding of the landscape alteration through time, subsequently using that training to make future predictions (Lambin 1997; Pijanowski et al. 2002). Further, we found a balance between deterministic (17) and stochastic (18) models, with a significant relationship between model type and stochascity ($\chi^2 = 21.27$, df = 4, $p < 0.001$). All optimisation models were found to be deterministic, whereas 55% of cellular automata were stochastic models.

3.4.3) Drivers of deforestation in the Amazon

Amazonian LULC change models used an average of 10 inputs, with some models using as few as five (Walker et al. 2004; Nepstad et al. 2009; Müller et al. 2011) and one as many as 40 (Moreira et al. 2009). Across papers, we found that model inputs fell into four broad categories; (1) geographical, (2) economic, (3) social, and (4) biological inputs (Fig. 9). Every model we investigated used a geographical input of some description, and typically used these inputs to aid in determining the spatial location of changes. The three geographical inputs that were used most consistently were roads (24/35 papers), soil factors (20/35) and landscape factors (20/35). Distance to roads, urban centres and past deforestation is typically negatively correlated with future deforestation, with higher deforestation occurring in close proximity to these locations (de Koning et al. 1999b; Soler et al. 2007; Mann et al. 2010; Maeda et al. 2011). The suitability of land for agriculture influences deforestation probabilities, with nutrient rich soils more likely to be deforested than nutrient poor soils (Etter et al. 2006b; Soler et al. 2007). Further, deforestation tends to occur on flat land at low elevation and is much less likely on slopes which are harder to farm (Müller et al. 2011).

Economic inputs such as the price of farm goods, the value of land and gross domestic product (GDP) were used in 19/35 models and are typically used to predict the amount, rather than the spatial location, of LULC changes (de Koning et al. 1999b). Given that the vast majority of LULC change is associated with development (*e.g*. agriculture and resource extraction), it is not surprising that economic indicators, such as agricultural goods prices, make good predictors of how people and/or governments are likely to alter the land use of an area. For instance, Soares-Filho et al. (2004) found that 71% of the variance in annual deforestation rates was explained by gross national product, although Ewers et al. (2008) used time-series analyses to demonstrate there is no statistical evidence that any economic variables, including per capita GDP, have systematically caused variation in deforestation rates.

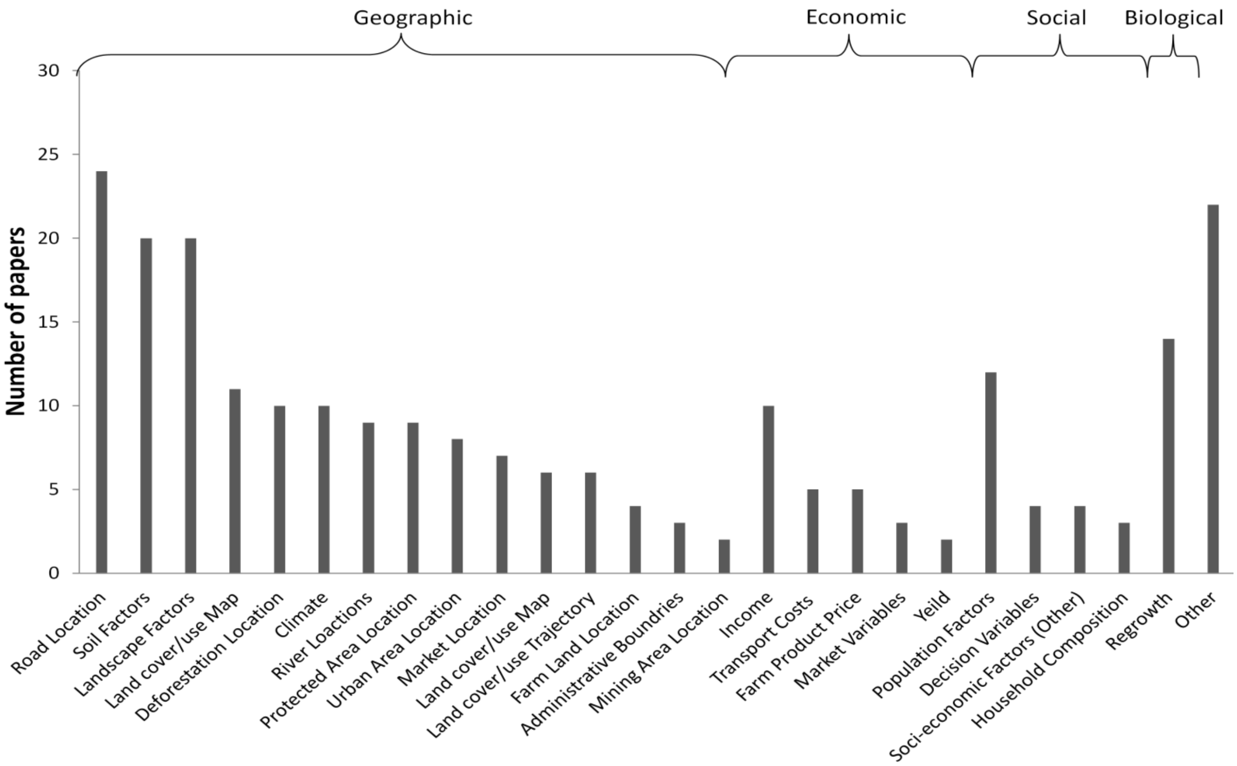


Figure 9 – The number of papers using each input type. Inputs are divided according to class; geographic, economic, social or biological. Soil factors have been put into a single group that consists of factors such as soil moisture and soil texture. Landscape factors include inputs such as altitude and slope; climate includes rainfall, temperature, dry season length etc. If a paper uses multiple inputs from a group it is still counted only once e.g. if soil fertility, moisture and texture are used it counts as one soil factor.

Almost half of the models (14/35) made use of social inputs to connect people to LULC change decisions based on assumptions about their behaviour. For example, Walker et al. (2004) showed that household demography was the main factor affecting land allocation (conversion) decisions. They suggested that a household economy framework, which takes into account social and economic factors, may be a more appropriate approach than simple profit maximisation approaches to LULC modelling (Walker et al. 2004). Nearly all LULC change over the last century has been a direct result of individual and social responses to changes in the economic climate (Lambin et al. 2003), and a key assumption of many economic-based models is that people will seek to maximise profit (Evans et al. 2001). This, however, may not be appropriate for the Amazon which represents a frontier setting, where the institutions necessary for profit maximisation may not be present or fully functional (Walker et al. 2004).

Research at both local and regional scales have found complex relationships, feedbacks and interactions between human (social, political, economic) and environmental systems (Deadman et al. 2004). One such relationship is that between road construction and deforestation, with this causal interaction driven by economic and cultural factors (Geist & Lambin 2002). Another common relationship is found between property rights and deforestation: Araujo et al. (2009) found that insecurity in property rights and social conflicts increased deforestation, because landowners needed to assert use of the land to avoid expropriation and squatters deforested in the hope that property rights will be awarded in the future. Differences in how models assume people will behave can exert large effects on model predictions, as shown by scenarios modelled by Dale et al. (1994) that compared alternative behaviours of farmers and their farming practices. In one scenario, it was assumed that farmers will make innovative use of their land and implement positive agro-forestry practices, leading to predictions that 40% of forested land would be cleared by farmers after 40 years. By contrast, when the model assumes that farmers will not use innovative practices and do not implement agro-forestry, the model predicted that 100% of the land would be deforested within just 10 years.

Finally, biological inputs included variables such as plant growth rates, agricultural yield and crop nutrient demands (*i.e.* the soil requirements of various crops). For example, crop nutrient demands in conjunction with soil fertility determines the viability of different crop types that might replace a forest, with highly fertile areas likely to become arable land (*e.g.* coffee or maize) and low fertility areas more likely to become pastoral land. Another biological input that was often used (13/35 models) was forest re-growth rate and/or the probability of forest re-growth (Soares-Filho et al. 2002). Distance to re-growth has also been used to predict deforestation, with the observation that deforestation and distance to re-growth are negatively correlated (Soares-Filho et al. 2002).

The relative importance of inputs varied with location, not only between models, but also within models working in different regions. For example, Wassenaar et al. (2007a) found that existing fragmentation was one of the most significant model inputs across seven Amazonian regions modelled, however there were regional differences in model structure. For instance, altitude was an important predictor of deforestation within the Ecuadorian Amazon, along the edge of the Andes mountain range, but was not important in the other six regions that were much less topographically complex. Also, Etter et al. (2006b) found that distance to towns and roads were important predictors of deforestation in both Andean and Amazonian regions, while soil fertility was important in the Andean but not Amazonian regions whereas the number of rain days was more important in the Amazon. These regional differences in the causes of deforestation patterns make it important that papers explicitly state the inputs they are modelling, but surprisingly this is not always the case. For example, Moreira et al. (2009) used '40 environmental, demographical, agrarian structure, technological and market connectivity indicators', but never listed them. Others mention the inputs but it is not always clear what they mean. For example Dale et al. (1994) and Soler et al. (2007) used 'soils' as an input variable, but do not specify if they are referring to soil type, soil fertility or soil texture. By contrast, de Koning et al. (1999b) explicitly stated that they used soil texture and fertility, finding that in the Andean region texture and soil fertility were both important modelling parameters, while in the Amazon region neither played a role at the scales modelled.

Landscape factors were typically static inputs to LULC models, although the LULC map itself represents an obvious exception, changing at each time step of a model as LULC change progresses (Messina & Walsh 2001; Soares-Filho et al. 2004; Walsh et al. 2008). Dynamic economic inputs were also observed, with each year's activities (conversion to farmland for instance) resulting in new stocks of finances and/or resources that become the foundation of the next year's activities (Carpentier et al. 2000). Not all dynamic inputs build through time as these examples above. For example,, Messina and Walsh (2001) used a cellular automata module to select locations for deforestation based on neighbourhood rules, implementing a random number generator to recreate the dispersiveness of deforestation and to allow for stochastic deforestation events, and Walker et al. (2004) determined the number of deforestation events through a probability model that used a uniform distribution. In both cases the use of a probabilistic or stochastic selection of deforestation events makes the amount and location of deforestation a dynamic input. Some inputs are actually dynamic but are treated as static in models, and this is particularly true of roads. Most models we examined used roads as an input, but of those more than two-thirds treated roads as a static input. Only papers based on the DINAMICA, IDRISI or LandShift modelling frame works used roads as a dynamic, spatially explicit phenomenon (Messina & Walsh 2001; Soares-Filho et al. 2004; Soares-Filho et al. 2006; Lapola et al. 2010).

3.4.4) Model calibration

In LULC change modelling there are two key aspects of LULC change that need to be estimated: the rate of change and the location of change. There are several calibration methods employed by the papers modelling LULC change in the Amazon, but no direct comparison of the different methods on the same datasets, making it difficult to quantify the relative reliability of the various options. All calibration techniques apply statistical techniques to empirical observations of historical data to estimate parameter values and weights (Appendix A). Some model calibrations were combined with expert knowledge to capture inputs known to be important despite a statistical model simplification process removing them (Soler et al. 2007) and one-off events such as changes to agricultural subsidies (Wassenaar et al. 2007a).

Economic approaches to modelling LULC change tend to use more process-based methods for calibrating models than do other techniques that rely more heavily on extrapolating spatial patterns. Some models developed a 'demand module' that estimated the economic demand for particular agricultural product and used that to determine the amount of land needed to be converted (de Koning et al. 1999b). Similarly, where a key aim is to maximise profit or minimise costs, calibration techniques such as linear programming can be used to derive model input values that give rise to optimal solutions. This approach was employed by Carpentier et al. (2000) and Labarta et al. (2008), both of whom used agent based, farm-level modelling where the main goal was to maximise household income.

3.4.5) The difficulties of model validation

We found that 16 models only validated a single year of predictions, and in four of those models the time period used in the validation was the same as used to calibrate the model, suggesting a degree of circularity in the validations (Soler et al. 2007; Wassenaar et al. 2007a; Lopez & Sierra 2010; Maeda et al. 2011). Just two out of 35 models were validated at two points in time (Carpentier et al. 2000; Soares-Filho et al. 2002), two models were validated at three points in time (Deadman et al. 2004; Silvestrini et al. 2011), and only one study validated their predictions at four points (Evans et al. 2001). Thirteen models did not clearly state a validation method (Laurance et al. 2001b; Ferraz et al. 2005; Sarkar et al. 2009), used just visual comparison (Moreira et al. 2009; Mann et al. 2010) or argued that the modelling approach had been validated elsewhere (Dale et al. 1994; de Koning et al. 1999a; Soares-Filho et al. 2006; Nepstad et al. 2009) (Table A1 in Appendix A).

It is self-obviously problematic to validate predictions for a future that has not yet happened; yet modellers still have the option of employing backward validation, which involves running a model in 'reverse' to predict historical rather than future land use patterns. For instance, de Koning et al. (1999b) modelled deforestation in Ecuador from 1991 to 2010 and validated their model by using it to backcast LULC changes from 1991 to 1974. This allowed them to validate their model against an extensive land-use dataset based on an agricultural census carried out in that year. They found a strong positive correlation between their model predictions and observed LULC patterns with correlation coefficients varying between 0.71 and 0.96.

3.4.6) Quantitative assessment of model performance

Out of 35 published models, we were able to collect just eleven data sets either directly from the authors or via downloadable content (Appendix A), highlighting a lack of utility of LULC models. We focus on the Brazilian Amazon for which we were able to obtain independent deforestation data (INPE 2012b) against which to validate the model predictions. In this region, we only had three usable sets of model outputs that we could validate (out of 22 models from the Brazilian Amazon), two of which comprised a complete time series of model predictions (Soares-Filho et al. 2006; Yanai et al. 2012) whereas the other comprised a single map of predicted LULC at the end of the model prediction period (Wassenaar et al. 2007a). Different model scenarios, such as the Business as Usual (BAU) and Governance (GOV) scenarios of Soares-Filho et al. (2006), and the Baseline (BS), with leakage (WL) and with reduced leakage (RL) scenarios of Yanai et al. (2012), were treated as separate models in our validations.

We calculated the three measures of prediction accuracy described above: (1) 'match',; (2) 'omission', and (3) 'commission'. When assessed year by year (annually), model match was very low (0 – 2%) for both of the Soares-Filho et al. (2006) scenarios as well as the (Yanai et al. 2012) scenarios (1 – 3%) (Fig. 10a and Table A2 in Appendix A). Cumulative match, which compares the accumulated deforestation patterns from the start of the model until a given time point in the future (2003-2010), was also low for Soares-Filho *et al.* (2006) scenarios (1 – 7%) but increased as model duration extended (Fig. 10b). A similar pattern was found for Yanai et al. (2012) where cumulative model predictions reached 4 – 5% by 2010 (Fig. 10b). Commission rates in Soares-Filho *et al.* (2006) predictions also increased through time both on annual (from 4 to 8%) and cumulative (from 4 to 6%) comparisons (Fig. 10a and b, respectively) whereas omission errors decreased through time for both cumulative and annual comparisons (Fig. 10a and b). Only one deforestation map was available for Wassenaar et al. (2007a) predictions, showing cumulative deforestation between 2000 and 2010, during which period the model had a total percentage match of 10%, commission of 59% and omission of 31% calculated from the map. The Wassenaar et al. (2007a) model tended to have higher omission errors in the eastern Amazon and higher commission errors in the south (Fig. A1 in Appendix A), whereas the Soares-Filho et al. (2006) models had highest commission errors in the north and relatively evenly distributed omission and commission errors along the Arc of Deforestation (Fig. A2 and A3 in Appendix A).

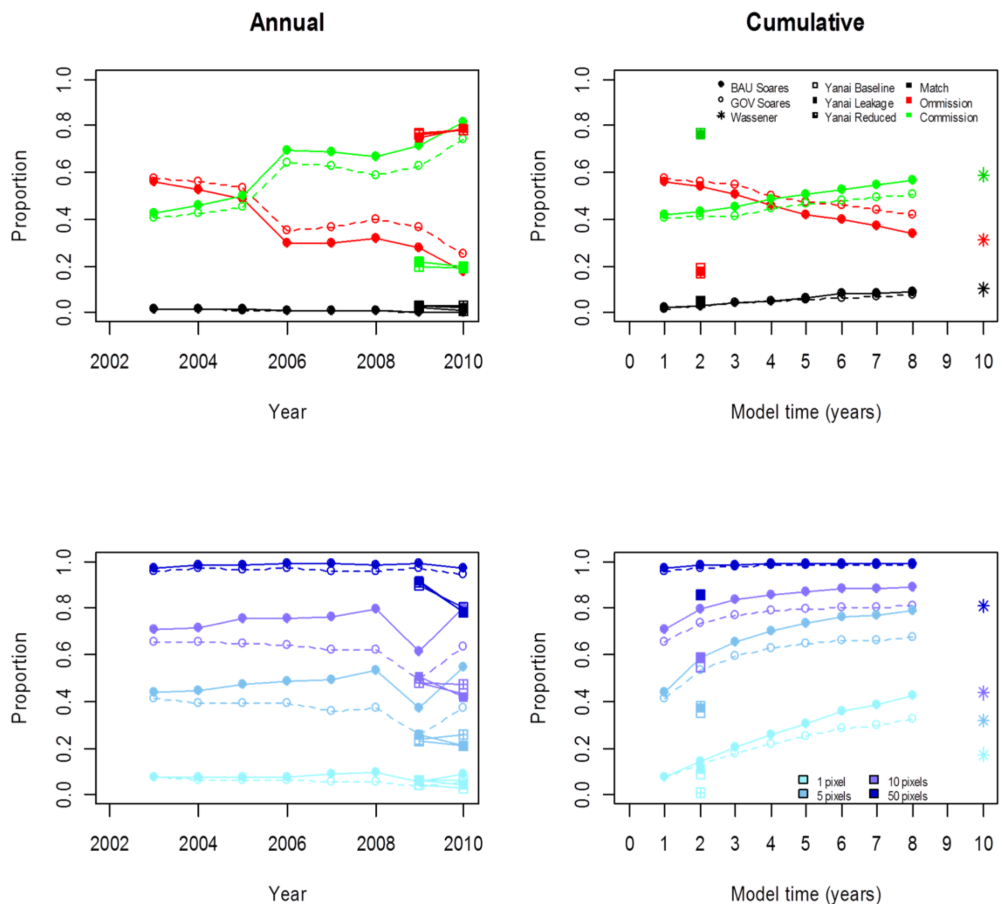


Figure 10 – Pixel-by-pixel comparisons between observed deforestation and predictions made (a) annually and (b) cumulatively by Soares-Filho et al. (2006) at 1 x 1 km grid cells, from both governance (GOV) and business-as-usual (BAU) scenarios between 2003 and 2010, and by Yanai *et al.* (2012) for 2009 and 2010 at 250 m grid cells for the baseline, with leakage and with reduced leakage scenarios. Proportion of observed deforestation within four distance classes (1, 5, 10 and 50 pixels) of predicted deforestation, calculated (c) annually and (d) cumulatively. Additionally, (b) and (d) show comparisons between observed deforestation and predictions made by Wassenaar *et al*. (2007) (5 km pixel size).

Given inherent difficulties in predicting the exact location of any spatial phenomenon, it can be misleading to assess model predictions based on a pixel by pixel comparison. As such, we refined our model validations by estimating 'how close' model predictions were to actual annual deforestation. We applied buffers of 1, 5, 10 and 50 pixels around the predicted deforestation and calculated the proportion of observed deforestation that fell within the various buffers (Fig. 10 c and d). For the Soares-Filho *et al.* (2006) models, annual predictions still had very low match rates at the smallest buffer size (1 km, which corresponds to 1 pixel), reaching just 6% (GOV) and 9% (BAU). Cumulative model predictions performed better, reaching a maximum match of 33% (GOV) and 42% (BAU) by 2010. Unsurprisingly, model predictions improved with increasing buffer size, with nearly 100% of all deforestation events falling within 50 km (or 50 pixels) of model predictions. The BAU scenario performed better in all comparisons which is likely because this scenario predicts higher overall rates of deforestation. A similar pattern was found for both Yanai et al. (2012) (250 m pixel size) and Wassenaar et al. (2007a) (5 km pixel size) model predictions, although for every buffer size these seemed to be consistently lower than Soares-Filho *et al.* (2006) predictions.

3.5) Discussion

3.5.1) Appropriate scales are process-specific

We found LULC models in the Amazon differed considerably in the spatial and temporal extent at which they are developed and applied. Small-scale models were usually developed at the farm or plot level (dozens to hundreds of km^2) (Dale et al. 1994; Deadman et al. 2004; Labarta et al. 2008), whereas medium-scale models were developed for sub-regions within a country or small countries (thousands of km^2) (de Koning et al. 1999b; Soares-Filho et al. 2004; Maeda et al. 2011) and large-scale models were built for larger countries or groups of countries (millions of km^2) (Soares-Filho et al. 2006; Wassenaar et al. 2007a; Lapola et al. 2010). Working at any of these scales has strengths and weaknesses. For example, farm-level models can simulate farmers' decisions and reactions to market variations such as changes in commodity prices. However, these models are usually site-specific, making it very difficult to generalise them to larger areas or other tropical regions. Larger scale models, on the other hand, tend to use aggregated data which often averages variability across the region modelled and therefore lose detail when interpreted at fine spatial scales (Mertens & Lambin 1997).

3.5.2) The biggest weakness in model formulation?

Despite knowing that road networks in the Amazon are highly dynamic (Brandão & Souza 2006), most models treat road networks as a static pattern (Appendix A: Modelling road expansion in the Amazon). Roads are the key spatial determinant of deforestation patterns (Forman & Alexander 1998; Fearnside 2005; Finer et al. 2008), determining the accessibility of land and cost of transportation which in turn determines the viability of land use change of a given area. Road maps, and distance to roads, were the most commonly used inputs for LULC change modelling based in the Amazon where the road network has expanded rapidly (Fearnside 2005; Brandão & Souza 2006; Ahmed et al. 2012), yet few models treat roads as a dynamic variable. We suggest the reason for this is that modelling the expansion of road networks is itself a formidable challenge and one that has been identified as a key weakness in our ability to predict LULC change in the Amazon (Barlow *et al*. 2011). Certainly there are several modelling frameworks available to predict the development of road networks and that were used in the LULC models included in our review (Messina & Walsh 2001; Soares-Filho et al. 2004; Soares-Filho et al. 2006; Lapola et al. 2010), but we were unable to find any peer-reviewed presentation of these road models, nor any numerical validations of the road model predictions. While it is clearly desirable to have a dynamic road model integrated with deforestation models, it is not so clear that an untested road model represents an improvement over the use of static road networks alone.

3.5.3) The way forward for validating LULC models

Although we recognise that the choice of number of years to validate model predictions is, in most situations, limited by the amount of data available, we suggest that the relatively limited attention paid to the act of model validation greatly limits the reliance that can be placed on LULC model predictions. It is self-obviously problematic to validate predictions for a future that has not yet happened, although back-casting provides one option. Even with back-casting validation, however, simple measures of correlation may not be enough to fully validate model predictions. As Pontius Jr et al. (2004a) pointed out, authors should also take into consideration land use classes that persist, increase and decrease in the landscape through the implementation of cross-tabulation matrices with quantitative disagreement and allocation disagreement measures (Pontius & Millones 2011). Further, measures of persistence, gains and losses of various land cover types are informative, but they can still fail to identify if these land cover changes are systematic (Pontius Jr et al. 2004a), in which case analysis of intensity may be more appropriate (Pontius Jr et al. 2004b; Huang et al. 2012). This is a method by which the losses and gains of a given land use category may be separately calculated to illustrate the direction of change, thereby accounting for the systematic nature of observed changes (Versace et al. 2008; Huang et al. 2012). This would allow the differentiation between random and systematic transitions which in turn would elucidate the model efficacy. We note, however, that none of these more detailed validation methods were applied in the Amazonian LULC models in our review.

Despite the many validation methods already available (Supporting Information), more theoretical work needs to be done on complexity theory and on emergent spatial patterns of LULC to develop more robust methods of answering the basic question of 'what is a good fit' for LULC change models (Messina & Walsh 2001; Messina et al. 2008). Despite the fact that some landscape level predictions can be verified by traditional tests of significance, the detailed spatial patterns that emerge from LULC change models remain largely unverified (Messina & Walsh 2001).

A further obstacle to statistically evaluating spatial model outputs is that there is disagreement upon baselines, or what represents an appropriate null hypothesis. For example, the widely used Kappa statistic simply estimates whether model outputs are better than those predicted by a random model (Pontius et al. 2007). Such an approach may not be informative because a model that predicts better than random does not necessarily mean it predicts well (Walker 2003). Perhaps a more appropriate question, raised by Pontius and Millones (2011), would be to assess how much 'less than perfect' the model is. The authors suggested the use of a naive model (i.e. alternative model) as a null hypothesis in comparisons. This baseline is built by using the model calibration quickly and naively to create a comparison map (Pontius & Millones 2011). For example, Wu et al. (2009) found that 89% of their calibration data was of a single land use category and, as such, defined a naive baseline predicting that their comparison map would all be of a single category. The authors then carried out an evaluation of these two different approaches (random *versus* naive) and reported that seven of eight models performed better than a random prediction but only one of eight models performed better than a naive model baseline. While the naive model approach reduces the number of 'successful' models, possibly because LULC change events are inherently non-random, it still fails to quantify how much 'less than perfect' the model predictions are. Although validation statistics need to be further developed, it is still informative to utilise simple measures of prediction accuracy such as the ones we present below, comparing predicted and observed LULC changes pixel by pixel.

Finally, we note that classification errors in the observed datasets can lead to undetected uncertainty around any estimate of model validation error. Given that these datasets are used as the 'truth' against which the model predictions are validated, uncertainty surrounding these reference maps needs to be minimal else all validations will be inherently flawed. This is also true for model calibration and all input variables used; if the data that underlies the model are poor, than the models will have poor predictive power. A systematic bias in observed LULC classification is likely to lead to biases in the model predictions and/or validation. By contrast, if errors in the LULC classification are random then the inaccuracies will lead to greater uncertainty around model predictions but not necessarily a biased prediction.

3.5.4) How certain are we about LULC model predictions?

We found remarkably low rates of prediction accuracy from the three LULC models we were able to test ourselves, and this was particularly true of year-by-year predictions. When accumulated over longer time periods, model accuracy invariably improved, suggesting that over long time frames it is possible to predict the spatial patterns of LULC change with moderate certainty. However, because of the inaccessibility of predictions from a larger set of LULC models, or our inability to obtain independent deforestation data against which to validate them, we have no way of determining if the rates of model success we quantified for Soares-Filho *et al.* (2006), Wassenaar *et al.* (2007) and Yanai *et al.* (2012) are typical of the field as a whole or unique to these particular models.

The program codes developed to run LULC change models are often not available (15 of 35 models used software that could potentially be used to replicate their predictions; IDRISI, DINAMICA, LandSHIFT, CLUE and TerraME), while others did not use commercially or freely available software and did not make the source code available. Had source codes been available, we would have been able to quantitatively compare model performance on a standardised data set, much as Elith *et al.* (2006) did to compare modelling approaches used to estimate species distributions. Still other papers did not provide full lists of their model inputs, again preventing us from replicating their results. We suggest that this failure to make model code and model predictions available to the wider community reflects and highlights a considerable shortcoming of the LULC change field as a whole.

It is particularly surprising that none of the models we reviewed presented any estimate of uncertainty around their model predictions. This is despite several models incorporating stochastic elements that allow for the uncertainty to be directly quantified. Unfortunately, the quantification of uncertainty is not yet a common practice, and we suggest that it represents an important step for the discipline. However, we note that even in a stochastic model, some one-off events are very difficult to incorporate, especially those that are related to human decisions. For example, new deforestation frontiers opening up in the north of Rondônia were not captured by deforestation models based on historical data (Soler et al. 2007), and it is unreasonable to ever expect them to predict this type of event.

3.5.5) The future of LULC change modelling

LULC change models have been prominent in the literature for many years, but our review has uncovered three key points that need addressing to improve the transparency, reliability and utility of LULC change models. These three issues have been raised individually in the past by various authors, and by drawing them together here we hope to stimulate improvements to the discipline. First, we have identified a lack of openness with regard to presenting and making available the model inputs, model code and model outputs that prevents the community from fully understanding and rigorously comparing models (Grimm et al. 2006). Second, there are considerable difficulties involved in validating model outputs, and indeed a lack of consensus on the appropriate techniques (Verburg & Veldkamp 2005; Messina et al. 2008). Third, there is no standardised model framework that can be used as a basis for comparing LULC models and generating multi-model inference (Grimm et al. 2006; Messina et al. 2008).

We suggest that the rise of climate change models provide a pathway that LULC change modellers may emulate to improve the discipline. While we recognise LULC modelling approaches differ substantially from each other, likely exhibiting more model-to-model variation than the modelling approaches used in climate change, the key aim of predicting LULC change is the same regardless of the approach taken (*i.e.* to predict the future of the landscape), thus model outputs should be made available for people to use and to compare. We certainly do not advocate a move towards producing a unified methodology where every model uses the same code and approach, recognising that there is strength in having a diversity of approaches. But we do believe that to harness that strength, we need to be able to overlap the spatial predictions of all models, weight those predictions by the validation errors of the models that generated them, and thereby gain informative among-model comparisons and provide a basis from which to make predictions based on model averaging.

Climate change modelling has made considerable progress towards solving some of the aforementioned problems. The World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project (CMIP3) set up an archive named 'WCRP CMIP3 Multi-Model Dataset' to provide IPCC model outputs in a standardised format (WCRP 2012). We believe that a similar archive for LULC change predictions would prove invaluable for further progression in the field. In line with this, model code and methodologies should be available in an easy to understand and accessible format to facilitate output comparisons, rather than remain proprietary software or unpublished as is the case with most existing LULC change models (Appendix A, Appendix A2). Model documentation in such an archive should be much more detailed than it is in many of the models published in the peer-reviewed literature, ensuring, as an obvious example, that at the very simplest level the full list of model inputs are named.

Since the early 1990s atmospheric climate modellers adopted a standard protocol for GCMs (General Circulation Models) (Gates 1992). The protocol provided a framework for model diagnosis, validation and inter-comparison (Tebaldi & Knutti 2007) and has since been used widely. The field of LULC change modelling would benefit from a similar framework, and particularly so to ensure high standards of model validation, a stage of the modelling process that is overlooked or poorly executed in the majority of models we examined. Agent-based modelling has made some headway toward this with the Overview, Design concepts, and Details (ODD) framework (Grimm et al. 2006), the MR POTATOEHEAD framework (Parker et al. 2008) and work done by Polhill and Gotts (2009). One advantage gained from standardised procedures is that they allow the analysis of multi-model ensembles, which are now commonly used in climate modelling and form important components of reports from the Intergovernmental Panel on Climate Change. Combining models in an ensemble increases the reliability and consistency of predictions (Tebaldi & Knutti 2007), and this approach has found utility in other disciplines such as public health (Thomson et al. 2006) and agriculture (Cantelaube & Terres 2005). Disparities among LULC models in terms of scale and resolution, combined with a more basic failure to make model predictions available, currently prevents any such methods from being applied to questions of LULC change. Furthermore, increasing the availability of source code would allow comparisons of models based on the same input data to be carried out as was done by Elith et al. (2006) for species distribution models, allowing fair comparisons to be made among modelling techniques.

Although the majority of papers we reviewed do clearly state how models were built and implemented, in some this is very difficult information to extract, particularly where calibration and validation methods are in question. This state of affairs clearly needs to be improved, and would not only inherently increase the transparency of models but also increase their utility, as any outputs that are made available have clear provenance. While we recognise it is time consuming and often difficult to make model code understandable and useable by others, the uploading of model outputs as rasters, .shp, ASCII and other widely used formats is not. This should become standard practice, yet currently it is difficult to obtain the outputs of published models, as testified by our ability to obtain just 11 out of 35 datasets (of which only three presented year-by-year model predictions as opposed to a single, final-year prediction). We suggest that all LULC change models that are published should provide, as a minimum, (1) a reference dataset of LULC at two time steps, one of which should be the beginning of the model simulation, and (2) the predicted dataset for each time step(s) in the simulations that can be used for comparative analyses.

3.6) Conclusions

Climate change models have exerted considerable influence over public perceptions of climate change and now impact policy decisions at all political levels. We suggest that LULC change models have an equally high potential to influence public opinion and impact the development of land use policies based on plausible future scenarios, but to do that requires a step-change in the discipline. Developing a set of standardised procedures for methodology, validation and output reporting would greatly increase the utility of LULC models, and will be a necessary step towards generating reliable scenarios that can be used to influence environmental policy at global scales.

3.7) Acknowledgements

The authors would like to thank the Grantham Institute for Climate Change, Microsoft Research Cambridge and European Research Council for funding. The authors would like to thank those authors who provided their spatial data for us to use: Soares-Filho, B., Lapola, DM, Wassenaar, T, Soler, LD, Maeda, EE, Messina, JP, Walsh, SJ, Etter, A, Vadez, V., Yanai, A, and Müller, R. We would also like to thank Carpentier, C., Dale, V.H., Deadman, P., Evans, T.P., Lapopa, D.M., Laurance, W., Maeda, E., Mann, M.L., Messina, J.P., Sangermano, F., Soares-Filho, B., Soler, L., Vadez, V., Walker, R., and Wassennar, T. for commenting on a draft of this manuscript.

Chapter 4 – Predictive modelling of contagious deforestation in the Brazilian Amazon

Isabel M.D. Rosa[1], Drew Purves[2], Carlos Souza Jr[3] and Robert M. Ewers[1]

[1]Imperial College of London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, United Kingdom

[2]Computational Ecology and Environmental Science, Microsoft Research Cambridge, Roger Needham Building, 7 J J Thomson Ave, Cambridge CB3 0FB, United Kingdom

[3]Imazon – Amazon Institute of People and the Environment, Rua Domingos Marreiros, 2020, Fátima, CEP: 66.060-160, Belém, Pará, Brazil

Manuscript submitted to *PLOS One* in September 2012

Revised manuscript accepted for publication in *PLOS One* in September 2013.

Keywords: spatial probabilistic model, Brazilian Amazon, land-cover change, conservation, contagious deforestation

Running title: predictive deforestation model Amazon

Corresponding author: Isabel MD Rosa (i.rosa09@imperial.ac.uk)

4.1) Abstract

Tropical forests are diminishing in extent due primarily to the rapid expansion of agriculture, but the future magnitude and geographical distribution of future tropical deforestation is uncertain. Here, we introduce a dynamic and spatially-explicit model of deforestation that predicts the potential magnitude and spatial pattern of Amazon deforestation. Our model differs from previous models in three ways: (1) it is probabilistic and quantifies uncertainty around predictions and parameters; (2) the overall deforestation rate emerges 'bottom up', as the sum of local-scale deforestation driven by local processes; (3) deforestation is contagious, such that local deforestation rate increases through time if adjacent locations are deforested.

For the scenarios evaluated – pre- and post-PPCDAM ("Plano de Ação para Proteção e Controle do Desmatamento na Amazônia") – the parameter estimates confirmed that forest near roads and already deforested areas is significantly more likely to be deforested in the near future and less likely in protected areas. Validation tests showed that our model correctly predicted the magnitude and spatial pattern of deforestation that accumulates over time, but that there is very high uncertainty surrounding the exact sequence in which pixels are deforested. The model predicts that under pre-PPCDAM (assuming no change in parameter values due to, for example, changes in government policy), annual deforestation rates would halve between 2050 compared to 2002, although this partly reflects reliance on a static map of the road network.

Consistent with other models, under the pre-PPCDAM scenario, states in the south and east of the Brazilian Amazon have a high predicted probability of losing nearly all forest outside of protected areas by 2050. This pattern is less strong in the post-PPCDAM scenario. Contagious spread along roads and through areas lacking formal protection could allow deforestation to reach the core, which is currently experiencing low deforestation rates due to its isolation.

4.2) Introduction

The Amazon is the largest remaining continuous tropical rainforest on Earth. It covers about 6 million square kilometres and crosses nine nations' boundaries. Brazil is the country that hosts the largest portion (about 60% of the area) of the Amazon. This region is characterized by its high cultural and biological diversity(Barlow et al. 2011), but by 2009 already 19% of its forest cover had been converted to other land uses(Pereira et al. 2010a). Deforestation models have been developed to predict which areas are more likely to be deforested in the future and to simulate the impacts of different conservation and market strategies(Laurance et al. 2001a; Soares-Filho et al. 2006), or climatic trajectories and environmental policies(Lapola et al. 2011), on the spatial patterns of future forest cover.

The rate of deforestation – that is, the area deforested per year – in the Brazilian Amazon is highly variable (Ewers et al. 2008). These fluctuations are related to several factors such as the economic health of the country, infrastructure development, and the world's demand for agricultural products, such as beef or soybeans(Fearnside 2005; Morton et al. 2006; Nepstad et al. 2006b; Ewers et al. 2008). More recently, governance through command and control, restriction to rural credits and expansion of protected areas, helped by a global economic crisis, seem to have contributed to reduce deforestation(Rosa et al. 2012) going in an opposite trend to Brazil's economic growth(Assunção et al. 2012).Although these regional and global factors influence the deforestation rates in the Brazilian Amazon, deforestation is ultimately the sum of thousands of local deforestation events, which occur with an intensity that varies greatly across the region due to many factors including physiographic attributes, access to infrastructure, human population characteristics and dynamics, and socioeconomic organization(Garcia et al. 2007).

Geist and Lambin (2002) identified two types of causes for tropical deforestation: proximate causes (*e.g.* infrastructure expansion and agriculture expansion) are human activities that directly lead to change at the local level; and underlying causes, which can be demographic (population dynamics), economic (economic growth or change), technological (improvement or development) or political (environmental laws or policies). When modelling land cover change, modellers aim to select statistically variables that best represent these causes at the scale the model is being developed. For example, economic variables in small-scale models might include the decisions of private actors such as farmers who decide whether they will deforest part of their land (Evans et al. 2001; Walker et al. 2004; Vadez et al. 2008).By contrast, larger-scale models (such as the one we present here), cannot address this fine-scale decision-making process and instead focus on what drives deforestation at the regional scale, such as landscape-scale changes in agricultural land and/or infrastructure development plans (Soares-Filho et al. 2006; Wassenaar et al. 2007a; Maeda et al. 2011). However a model may represent the process of deforestation, and whichever predictor variables it may include, it is crucial that the models are constrained against observational data, such that their predictors are at least consistent with the rates and patterns of deforestation observed in the recent past.

The single most important factor that drives deforestation in the Brazilian Amazon is agricultural expansion. Climate and soils are the main constraints to agriculture(Sombroek 2001), and infrastructure such as roads determine the ease with which agricultural products can be transported to market. The inaccessibility and humid climate of the northwest regions of the Amazon leaves states like Amazonas less agriculture-prone, whereas along the southern and eastern margins of the Amazon, favourable climates for agriculture combined with extensive road networks explain the concentration of deforestation activity in this particular region of the Amazon(Chomitz & Thomas 2003).

To move beyond these generalizations and to develop policies for managing the balance between agriculture, biodiversity, and carbon storage in the Amazon, requires models that can predict the potential magnitude, timing and spatial patterns of deforestation under different policy scenarios(Wearn et al. 2012). Several such models, focussing on different aspects of the problem, have been published(Laurance et al. 2001a; Soares-Filho et al. 2006; Wassenaar et al. 2007a; Moreira et al. 2009; Lapola et al. 2011). As expected, there are still high uncertainties attached to projecting the location and rate of future deforestation(Achard et al. 2004; Houghton 2005) intrinsic to any modelling methodology, mostly because deforestation is statistically rare in the Brazilian Amazon (i.e., the large majority of the forest areas remain unchanged). In addition,, uncertainties in predictions arise from differences in the proximate and underlying processes that the models attempt to replicate, and limitation of data in adequate time and space scales. Several models predict the potential spatial pattern of forest cover in the Brazilian Amazon under scenarios that maintain the overall deforestation rate as it is today (business as usual) or scenarios that assume implementation of additional government measures (governance), either for the whole region (Laurance et al. 2001a; Soares-Filho et al. 2006) or for sub-regions Soares-Filho et al. (2006). Additional studies have used more explicit policy scenarios, such as changes in law enforcement (Moreira et al. 2009), or climate change (Lapola et al. 2011),to adjust the regional deforestation rates that drive the models.

Deforestation models use a set of biophysical and socio-economic variables, such as accessibility maps (mainly roads and rivers), landscape maps (land-cover/land-use), cattle and soy prices, human population density and agricultural suitability maps, to predict where deforestation is more likely to occur in the future. Although using different methodologies, they all agree that maintaining the rate of deforestation at current levels would have devastating impacts on the ecosystem and atmosphere, and agree about the relative risk among different regions. Laurance et al. (2001a)used a simple spatial model to generate two scenarios for the future of the Amazon, with the main difference being the effectiveness of protected areas in preventing deforestation. Both scenarios suggested a dramatic landscape alteration, ranging from 28% to 42% of the region deforested or heavily degraded over the 20 year period beginning in 2001, especially in the south-eastern areas of the Brazilian Amazon. The authors concluded that the efforts to avoid deforestation by improving conservation will be overwhelmed by the destructive trends observed in this region. Soares-Filho et al. (2006), although using an improved methodology that allowed for different deforestation rates among the 47sub-regions of the Amazon, found a similar effect, albeit one that took an additional three decades to manifest. All these policy-sensitive scenarios revealed that, given a regional deforestation rate, the spatial pattern will continue to be mostly concentrated in the eastern part of the Amazon where the infrastructures are well developed. Similar results were found by Wassenaar et al. (2007a), who used the modelling environment CLUE-S to model deforestation in Central and tropical South America until 2010.

Here, we introduce a dynamic and spatially-explicit predictive model of deforestation for the Brazilian Amazon. Our model captures three important aspects of deforestation:

uncertainty, emergence, and contagion. The first source of uncertainty is due to deforestation, at the local level, being probabilistic. Because of this stochasticity, we could not expect to predict the details of deforestation perfectly, even if we could predict the magnitude and regional spatial pattern perfectly (by analogy, we could not be expected to predict whether a coin would land heads up or tails up, even if we knew that it was fair). In addition, there is uncertainty in the model structure (*e.g.* best set of predictor variables to use, and how to include their effects), the values of predictor variables (*e.g.* they might be derived, with some error, from satellite images), and in the parameter values of models (*e.g.* the coefficient that determines the impact of a given predictor variable on the probability of deforestation). As such, models predicting deforestation should allow for the calculation of uncertainty on the predicted magnitude, timing and spatial patterns of deforestation resulting from the inherent stochasticity of deforestation events (Lewis et al. 2009a). Although many models in the literature do include stochastic elements in their approaches (e.g. [4,5]), they do not take advantage of this to provide spatial uncertainty measures associated with their outputs; the uncertainty is only provided by the means of different scenarios. The uncertainty associated with model predictions is crucial to policy makers who need to weigh up the model predictions against other considerations, and other models.

Emergence refers to fact that regional or country-wide deforestation rates (or amount of forest loss) are the sum of deforestation occurring at the local scale, influenced by local factors (such as proximity to roads), and local processes. Even regional or global drivers occur via local processes (*e.g.* changes in tax regimes or law enforcement). Because of emergence, the local, and then overall, rates of deforestation can change through time in ways that are not readily anticipated when viewing the phenomenon at the regional scale. Emergence of new deforestation is modelled stochastically but it is driven by local social-economic drivers.

In contrast to previous models, simulating deforestation as an emergent phenomenon allows our model to predict how the overall deforestation rate (or the rate in different regions) might change as deforestation moves across a landscape. This is a fundamentally different approach from accepting the overall rate as a top down input that is imposed upon a spatial model, with that pre-determined amount of deforestation then distributed across the region.

This latter approach is widely used by many modelers (e.g.*[4,5,16]*). However, the advantage to our bottom-up approach is that we need to parameterise just one model rather than two: both the top-down and bottom-up approaches need to parameterise spatial models to distribute deforestation across a landscape, but the top-down approach requires a second, separate model to be parameterised to determine how much deforestation will occur.

Contagion refers to the fact that location surrounded by recently deforested land, are likely to be more likely to suffer deforestation themselves. This deforestation then increases the probability of other nearby locations. Capturing contagion is crucial because it allows deforestation to spread through space. There is an analogy here with epidemiology(Murray 2002; Funk et al. 2010): once a disease [deforestation] first invades [begins] in a local region where there are susceptible individuals [forest], it can spread rapidly, especially if there a vector of transport or easy access [roads, rivers, etc.] between infected hosts [deforested areas] and susceptible individuals [forest areas].

Some models in the literature have made use of a 'patch expander' function and cellular automata models (*e.g.(Soares-Filho et al. 2006)*), which are based on neighbourhood effects and have some similarities to contagion. However, the 'patch expander' function is only used post-probability analysis: when calculating the weights of the variables to determine the probability of deforestation, the neighbourhood is accounted for, but as a distance metric (distance to previous deforestation) (*e.g.(Soares-Filho et al. 2006)*). Following a 'seed' deforestation event, the 'patch expander' function is used to create a spatial arrangement that more closely approximates reality, but it depends on pre-determined spatial probabilities of deforestation, rather than influencing those probabilities itself. Further, given that the rate of change is imposed 'top-down' in these models, the neighbourhood effect can only influence the location of change, but not the rate of change. By contrast, in our model we embed the neighbourhood effect into the model itself, allowing this contagious process to influence where, and also how much, deforestation will occur. Contagion is also important in the way it combines with stochasticity/uncertainty, because it allows random deforestation to spread into local clusters of deforestation, leading to patterns of deforestation that are very different from those that come from applying deforestation homogenously within regions. Finally, we improve on previous models by conducting a series of stringent model validation tests, comparing the model predictions of one scenario with observed deforestation events over a nine year period.

4.3) Materials and methods

4.3.1) Data sources

The first step of the modelling procedure was to identify the main drivers of deforestation in the Brazilian Amazon. The qualitative conclusions from a large literature are that deforestation occurs primarily near previously deforested areas(Alves 2002; Aguiar et al. 2007), near roads(Angelsen & Kaimowitz 1999; Alves 2002; Laurance et al. 2002; Pfaff et al. 2007; Asner et al. 2009), near markets(Pfaff 1999; Alves 2002; Aguiar et al. 2007),in areas with a pronounced dry-season(Laurance et al. 2002; Aguiar et al. 2007), and in regions that have been previously logged(Asner et al. 2006). Deforestation does occur in protected areas, but the rate tends to be lower than outside protected areas, as is the case in Rondônia (Nepstad et al. 2006a; Aguiar et al. 2007; Asner et al. 2009; Soares-Filho et al. 2010). However, in many other parts of the Brazilian Amazon this apparent effect is partially confounded with the fact that protected status tends to be conferred on relatively isolated regions where rates would be expected to be lower anyway (Joppa et al. 2008). The likely underlying causes of deforestation in the region are changes in gross domestic product (GDP), agricultural GDP, the size of the live cattle herd, and the rate at which temporary and permanent agriculture are expanding (Ewers et al. 2008; Ahmed et al. 2013).

We obtained input data to represent these proximate and underlying causes of deforestation (Table 1). The data was mostly obtained from three Brazilian institutions: Brazilian National Institute for Space Research (INPE), Brazilian Institute for Geography and Statistics (IBGE) and Amazon Institute of People and the Environment (Imazon). It included maps of historical deforestation, forest cover, road networks (official and unofficial), protected areas, rivers, topography, settlements and soil fertility. Dry season length maps were created by applying the methodology developed bySombroek (2001) to the historical monthly precipitation data (1960–1990) in the Brazilian Amazon, obtained from the World

Meteorological Organization (WMO). Economic data for each of the ~700 municipalities of the Brazilian Amazon were obtained from IBGE, representing municipality GDP and agricultural GDP, the size of the live cattle herd, and the land area under temporary and permanent agriculture.

Table 1 – Details of the input data used to calibrate the model for the transition period 2001-2002 and 2009-2010(data name, description, source, reference year and type)

Name	Description	Source	Year	Type
Deforestation	Annual deforestation	INPE[1]	2002 and 2010	Polygon
Previous deforestation	All deforestation occurred	INPE[1]	Until 2001 and 2009	Polygon
Forest cover	Remaining forest cover	INPE[1]	2001 and 2009	Polygon
Roads	Only main rivers	IMAZON[2]	2004 and 2007	Polyline
Rivers	Official and unofficial roads	IBGE[3]	-	Polyline
Settlements	Includes main cities, villages, and smaller settlements	IBGE[3]	-	Points
Topography	Altitude in km	SRTM[4]	-	Raster
Protected areas	Include indigenous lands, federal and state reserves	IMAZON[2]	2001 and 2009	Polygon
Soil Fertility	Reclassified for three classes: low, medium and high	IBGE[3]	-	Polygon
Dry Season Length	Number of months with precipitation < 100 mm	WMO[5]	1960/90	Points
Live Cattle	Number of head per municipality (heads)	IBGE[3]	2001/02 and 2009/10	Converted polygon
Temporary Agriculture Area	Total area of temporary agriculture (ha)	IBGE[3]	2001/02 and 2009/10	Converted polygon
Permanent Agriculture Area	Total area of permanent agriculture (ha)	IBGE[3]	2001/02 and 2009/10	Converted polygon
Gross domestic product	Municipalities' gross domestic product	IBGE[3]	2001/02 and 2009/10	Converted polygon
Agricultural gross domestic product	Municipalities' agricultural gross domestic product	IBGE[3]	2001/02 and 2009/10	Converted polygon

1 INPE – Instituto Nacional de Pesquisas Espaciais (http://www.dpi.inpe.br/prodesdigital/prodes.php)
2 IMAZON – Instituto do Homem e Meio Ambiente da Amazônia (http://www.imazon.org.br/)
3 IBGE – Instituto Brasileiro de Geografia e Estatstica (http://www.ibge.gov.br/home/download/geociencias.shtm)
4 SRTM – The Shuttle Radar Topography Mission from National Aeronautics and Space Administration (NASA http://www2.jpl.nasa.gov/srtm/)
5 WMO – World Meteorological Organization (http://www.agteca.com/climate.htm)

The data were separated into two categories of variables: static and dynamic variables(Soares-Filho et al. 2002). Static variables represented features that are assumed to stay constant through time such as soil fertility, topography, main rivers, state and dry season length. In our model, we also assumed that the distribution of protected areas is a static feature of the region, although the last decade experienced rapid and large expansion of protected areas. Dynamic variables, by contrast, represent features that change through time. In the model we present here, only forest cover itself and the proportion of deforested neighbour cells were treated as dynamic. There are several variables that are dynamic but which we considered to be static in our model due to a lack of data. This includes the economic variables (GDP, cattle herd, area of temporary and permanent agriculture), as we do not have economic models available to predict the value these variables take in the future. Similarly, we considered the road network to be a static feature of the landscape. Note that road networks in the Brazilian Amazon are known to be expanding (Brandão & Souza 2006), suggesting they should be considered as dynamic rather than static variables. However, in the absence of a validated model of road network expansion, we were unable to replicate this process and hence treated road networks as a static landscape feature (see Discussion). Static variables were calculated just once, in the beginning of the modelling process, whereas dynamic variables were re-calculated at each time step (year). Model variables were estimated individually for 5 x 5 km pixels across the Brazilian Amazon.

In contrast to previous modelling, we used as metric of local deforestation the proportion of deforested grid-cells in the neighbourhood of the focal cell. This contrasts with the usual approach of using Euclidean distance to the closest deforested cell(Laurance et al. 2001a; Soares-Filho et al. 2006; Lapola et al. 2011). We made this decision because our model updates the local deforestation probabilities as the neighbourhoods change through time. Within this dynamic framework, the Euclidean distance metric results in a very rapid, but diffuse, expansion of deforestation, characterized by rapidly spreading regions, within which there are a few deforested cells within a matrix of intact forest. This diffuse pattern of deforestation is not apparent in current deforested landscapes. The rapid expansion in models occurs because deforestation in a single cell immediately reduces the Euclidean distance over a large neighbourhood around that cell (in fact all cells, anywhere in the whole region, which are closer to the new deforestation event than to any previous event, experience an increase in deforestation probability). The diffuse pattern occurs because, if a cell already has a single deforested cell nearby, further deforestation events can have no effect on the local rate of deforestation (any event further away than the closest previous event has no further effect on local deforestation). By contrast, when using neighbourhood metric such as employed here, the local deforestation probability responds only to neighbouring cells, and builds continuously as the surrounding neighbourhood is deforested. As a result, simulations using the neighbourhood metric result in deforestation spreading in a well defined front, characterized in space as a rapid gradient from intact forest, to almost pure deforestation – a pattern that is consistent with observed patterns of deforested land.

4.3.2) Model structure, parameterization, and selection

Our model is based around $P_{defor,x,t}$, the probability that cell x becomes deforested in a set interval of time t. The fact that $P_{defor,x,t}$ is specific to a given time t illustrates how our model updates the local deforestation through time. We defined this probability as a logistic function:

$$Pdefor_{x,t} = \frac{1}{(1+\exp^{-k_{x,t}})} \tag{1}$$

such that as $k_{x,t}$ goes from minus infinity to plus infinity, $P_{defor,x,t}$ goes from 0 to 1. We could then write simple linear models for $k_{x,t}$ as a function of the driver variables affecting location x at time t. Similar logistic regression techniques have been successfully used and have become the standard method for assessing deforestation probabilities (Ludeke et al. 1990; Soares-Filho et al. 2002; Aguiar et al. 2007). The modelling procedure flowchart is shown in Figure 11 and the full model C^{++} code is provided in online (Supporting Information S1).

In 2004, the Brazilian government implemented the "Plano de Ação para Proteção e Controle do Desmatamento na Amazônia" (PPCDAM), which implemented a set of enforcement measures to fight illegal deforestation(IPAM 2012). The PPCDAM also coincided with a global economic downturn that reduced the economic incentives for deforestation(Rosa et al. 2012), and thus the post-2004 period represents a deforestation 'regime' that is very different to what was observed before. Before 2004, deforestation rates were increasing rapidly and deforestation was spreading quickly, whereas after 2004 the situation changed and the rates slowed down significantly(Rosa et al. 2012). Therefore, we fitted models to the data for observed deforestation that occurred pre-PPCDAM, between 2001 and 2002, whereas for a post-PPCDAM scenario the transition year used was 2009-2010.These two time periods represent very different deforestation regimes in the region, and thus the two calibrations of the model should result in different scenarios of deforestation by 2050.

We fitted 106 models to the observed deforestation patterns in each of the two transitions periods representing our two deforestation scenarios. Models differed only in the combination of static and dynamic variables included in the definition of $k_{x,t}$. To fit each of these 106 models we used 'Filzbach', a freely available C library developed by DP and others (http://research.microsoft.com/en-us/projects/filzbach/). Filzbach uses Markov Chain Monte Carlo (MCMC) sampling techniques to return, for each parameter, a posterior probability distribution from which we can extract the posterior mean, and a credible interval, given the model structure and the data. To carry out the parameter estimation, all that is necessary is to define the log-likelihood, which is a measure of goodness of fit between the model predictions and the observations, given a particular combination of parameters ?:

$$\ell(X \mid s,\theta) = \sum_{x,t} \log\{Z_{x,t} Pdefor_{x,t} + (1 - Z_{x,t})(1 - Pdefor_{x,t})\} \qquad (2)$$

where $Z_{x,t}$ is the observed deforestation at location *x* at time *t*, and *s* refers to one of the 106 models that we considered. The likelihood defined in Eq. 2 assumes independence among the samples, and is the same likelihood that underlies any standard logistic regression.

The exact set of 106 models was arrived at by forward stepwise regression. We chose the forward stepwise method, widely used in predictive studies(Davies et al. 2006; Aguiar et al. 2007; Wassenaar et al. 2007b) to first assess the impact of each variable on the probability of deforestation individually, and then to determine the additional predictive power gained by adding additional variables. In the first step, only the intercept is included, giving a one-parameter model. To assess this model, we used cross-validation, a statistical technique used to assess how accurately the model will predict data that was not used to train the model (Hastie et al. 2009). The cross validation was carried out by parameterising the model against a randomly selected subset of 50% of locations, then calculating the likelihood of the remaining 50% of the locations, using eq. 2. The purpose of the cross validation was to find a model that included those only predictor variables that had demonstrable predictive ability. Cross validation, where possible, is superior to model selection using information criteria such as the AIC, which is known to often lead to over-fitting (Liu et al. 2012; Wu et al. 2013). Next, each of the nine variables were added individually to the intercept-only model, creating a set of 2-parameter models that were again assessed with cross-validation. When all two-parameter models were trained and tested, the variable responsible for the highest maximum likelihood model was kept in the model and the remaining eight variables were again added individually. This procedure was repeated until all models were trained and tested, which resulted in a table with each model and the corresponding training and testing likelihoods, from which we selected the 'best model' as the one with the maximum test likelihood from all of the 106 models. However, after this procedure was complete, we found that some of the variables included in the best model had a non-significant confidence interval. In these cases, we chose the second best model, which had a slightly lower maximum likelihood, where all variables were in fact significant. This last, conservative, step, was carried out to further reduce the potential for over-fitting. The forward stepwise procedure was repeated for each of the two transition periods corresponding to the pre- and post-PPCDAM scenarios.

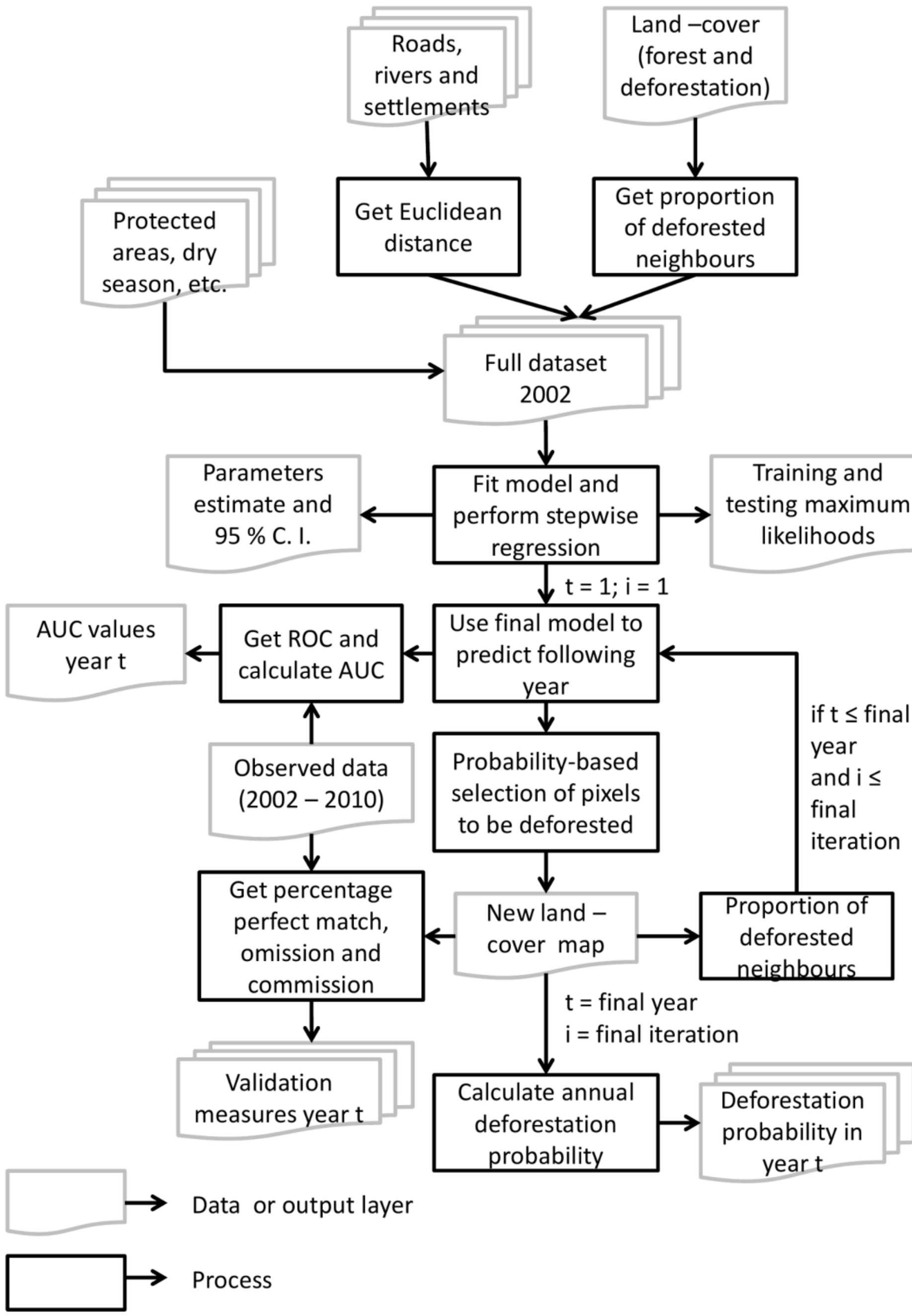


Figure 11 – Modelling procedure flowchart. The flowchart illustrates the construction and running of the deforestation model. *i* is the model iteration, *t* is the year, ROC refers to the Receiver Operating Characteristic and AUC is the area under the ROC curve.

4.3.3) Simulations

Once we had settled on the best statistical model for explaining past deforestation during each of the two transition periods, we used it to simulate future deforestation up to 2050 under the pre- and post-PPCDAM scenarios. To do so, all that was necessary was to re-apply eq. 1 in each time step, recalculating the dynamic variables (e.g. fractional deforestation around each location *x*), and using a slightly different set of parameter values at each iteration (to incorporate parameter uncertainty), drawn from a Gaussian distribution using the estimated mean and standard deviation for each parameter. This provided an updated $P_{defor,x,t}$ for each location *x*, which was then deforested with that probability. In practise, this was implemented as follows: for each *x*, draw a random number from a uniform distribution bounded at 0 and 1, and deforest *x* if this number is less than $P_{defor,x,t}$. After these deforestation events were implemented, $P_{defor,x,t}$ was calculated for every location *x* again, allowing for another round of deforestation. This procedure illustrates the three key aspects of our model mentioned above (see Introduction). The model is stochastic, because each individual deforestation event is drawn randomly using a weighted probability. Deforestation is contagious, because deforestation at location *x* increases the probability of deforestation at neighbouring locations, inducing further deforestation events which themselves further increase the probability of deforestation in the neighbours of the neighbours. Finally, the total (region-wide) deforestation rate at any time *t*, emerges as the sum of the local, stochastically determined, deforestation events, rather than being imposed top-down. The total deforestation rate can also vary through time, due to changes in the spatial configuration of forest cover in relation to the static and dynamic variables incorporated in the model. During each simulation, we kept a record of the fraction of cells undergoing deforestation at each time step, and a record of the pattern of forest vs. non-forest at each time step.

To calculate the uncertainty in model predictions, we ran the simulations multiple times (N = 100 iterations) and summarised the outputs across models at each time step. This allowed us to construct confidence intervals around our model predictions, rather than providing a single 'answer'. The uncertainty is represented by our final simulation outputs are a deforestation probability map calculated for each year in the simulation as the number of times a pixel was selected to be deforested in that year, divided by the total number of iterations. For each pixel that was deforested in the simulations, we also estimate the mean date at which it was deforested as well as the inter-quartile range around that date. This quantifies the uncertainty in the exact timing of deforestation events in the model simulations.

4.3.4.) Model validation

We validated our model predictions for the pre-PPCDAM scenario (parameterised with the transition year 2001-2002) against observed data for each year within the period 2002–2010 by calculating the area under the Receiver Operating Characteristic (ROC) curve in each year, which is used in many land-cover change studies (Etter et al. 2006a; Soler et al. 2007; Wassenaar et al. 2007a). For the post-PPCDAM scenario (parameterised with the transition year 2009-2010), the validation was only done for the first year of predictions (2010), reflecting the time period of deforestation data used to calibrate out model. For each of the 100 model iterations, we calculated the area under the ROC curve (AUC) value and three measures of precision on a pixel-by-pixel comparison: perfect match (the model predicts the exact location of deforestation), commission (over-predicting, the model predicts deforestation events that did not happen) and omission (under-predicting, the model did not predict deforestation in a location where deforestation happened). These three measures were calculated annually, using the observed and predicted annual deforestation maps; and for the pre-PPCDAM scenario were also calculated cumulatively, using the observed and predicted sum of deforestation at each time step (2002, 2002-2003, 2002-2004, etc.), for the time period from 2002 through 2010. Additionally, we calculated the proportion of observed annual and cumulative deforestation that occurred within certain distances (0, 5, 10, 25 and 50km) of our predicted deforestation at each time step.

4.4) Results

4.4.1) Model calibration

In both scenarios, the test likelihood returned from the cross validation increased rapidly with the addition of the first parameter, with additional parameters having progressively smaller impact on predictive power (Fig.12). Most variables were found to have the impact we expected on deforestation probabilities (Table 2). For instance, for both periods, distance to roads or rivers or settlements had a negative sign suggesting higher deforestation closer to these features. Also, protected areas had a negative sign, but here representing a lower probability of deforestation inside these areas when compared to unprotected land.

Further, annual increases in GDP, the size of the live cattle herd, and the area of land in permanent agriculture were found to have a positive impact on the probability of deforestation. By contrast, change in temporary agriculture area was non-significant in both cases, whereas change in agricultural GDP was found to only be significant in the period post-PPCDAM.

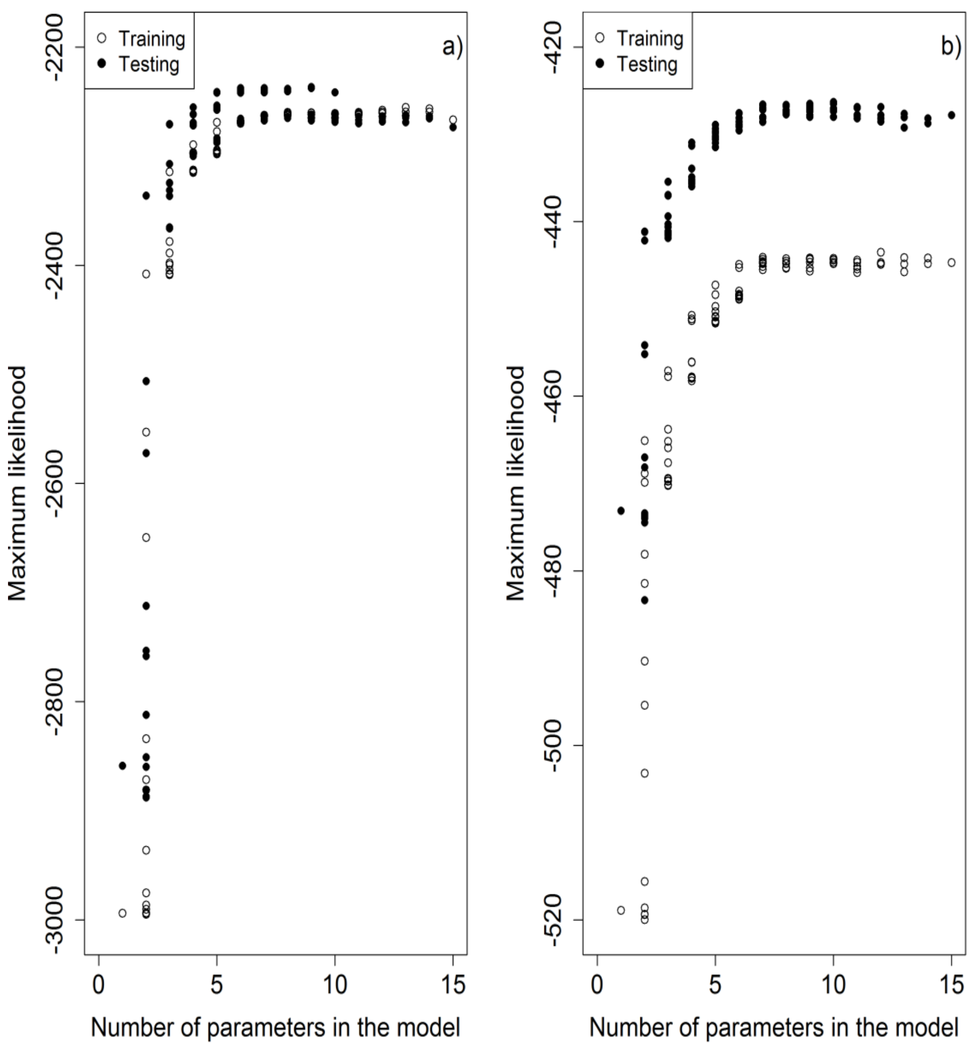


Figure 12 – Stepwise regression output. Figure shows both training and testing maximum likelihoods achieved by each of the 106 models used to explain deforestation rates in the Brazilian Amazon in the (a) pre-PPCDAM and (b) post-PPCDAM scenarios.

The most important variable was distance to roads, followed in turn by neighbourhood deforestation and protected areas, with our analysis indicating that deforestation probabilities were lowest in Indigenous lands, slightly higher in Federal and then State reserves, and highest in unprotected land. State also exerted considerable influence on deforestation probabilities and was retained in our best model. The only difference between scenarios, apart from variations in the effect size of individual variables (Table 3), was the inclusion of total GDP in the post-PPCDAM scenario. Adding additional variables had negligible effects on the test likelihood (Fig. 12) and were consequently omitted from the final model. The final set of parameter values used in the deforestation simulation for each scenario are shown in Table 3.

Table 2 – Mean and 95% confidence intervals of the single variable models, for each scenario (pre- and post-PPCDAM).

Parameter name	Pre-PPCDAM Mean	Lower limit	Upper limit	Post-PPCDAM Mean	Lower limit	Upper limit
Intercept	-4.7911	-4.8830	-4.6983	-5.9828	-5.9997	-5.9372
Previous Deforestation	4.9659	4.5774	5.2429	2.1113	-0.8230	3.2692
Roads	-0.00018	-0.00021	-0.00016	-0.00011	-0.00018	-0.0001
Rivers	-0.00002	-0.00003	-0.00001	-0.00005	-0.00007	-0.00003
Settlements	-0.00005	-0.00005	-0.00004	-0.00006	-0.00009	-0.00004
Topography	-1.9345	-1.9991	-1.7693	-1.8200	-1.9969	-1.5012
Soil fertility	0.0001	0.0000	0.0005	0.0002	-0.0002	0.0005
Dry season length	0.3194	0.2931	0.3394	-0.3075	-0.4019	-0.2190
Cattle	0.3947	0.1890	0.5294	0.1388	0.0013	0.3916
GDP	0.0532	0.0007	0.1535	0.0182	0.0005	0.0566
GDP agro	0.0064	-0.0002	0.0201	0.0447	0.0013	0.1882
Temporary Agriculture	0.0273	-0.0001	0.0849	0.0119	-0.0004	0.0413
Permanent Agriculture	0.3687	0.1534	0.5828	0.0658	0.0006	0.1966
Prot. Areas–State	-0.6281	-1.1474	-0.1856	-1.7402	-1.9976	-1.2311
Prot. Areas–Federal	-1.0426	-1.7319	-0.6574	-0.5949	-1.4540	0.0360
Prot.Areas–Indigenous	-0.9151	-1.1168	-0.7429	-0.3833	-0.8982	-0.1657
State – Maranhao	-0.4757	-1.2139	0.2250	0.3687	-1.2280	1.6955
State – Tocantins	-0.4989	-1.6543	-0.0464	0.2303	-1.6190	1.0590
State – Pará	-0.5673	-0.7502	-0.3772	-0.6250	-1.1601	-0.2796
State – Roraima	-0.9969	-1.3064	-0.7918	-0.3807	-1.0302	0.0041
State – Amapá	-0.8455	-1.3031	-0.5581	-1.1259	-1.9959	-0.3580
State – Acre	-0.3752	-0.5039	-0.2386	-1.1043	-1.8357	-0.4237
State – Rondônia	-0.1172	-0.2069	-0.0410	-0.3711	-1.0441	-0.0691
State – Amazonas	-0.5297	-0.6677	-0.4354	-0.4841	-1.0160	-0.2883
State – MatoGrosso	-0.0590	-0.1149	-0.0005	-0.3109	-1.1723	-0.0439

Table 3 – Mean and 95% confidence intervals of the final set of parameter inputs used in the deforestation simulations, for each scenario (pre- and post-PPCDAM). At each iteration, a slightly different set of parameters' values is drawn from these distributions to be used in the model that predicts deforestation from 2002 (or 2010 in the post-PPCDAM scenario) to 2050.

	Pre-PPCDAM			Post-PPCDAM		
Parameter name	**Mean**	**Lower limit**	**Upper limit**	**Mean**	**Lower limit**	**Upper limit**
Intercept	-3.70	-4.88	-2.57	-4.96	-5.92	-3.28
Previous Deforestation	2.10	1.70	2.57	2.32	1.61	3.07
Roads	-0.00011	-0.00013	-0.00009	-0.00006	-0.00009	-0.00003
GDP	-	-	-	0.45	0.08	0.76
Prot. Areas–State	-0.18	-0.73	0.09	-0.18	-0.97	0.52
Prot. Areas–Federal	-0.40	-0.84	-0.19	-0.71	-1.25	-0.19
Prot.Areas–Indigenous	-0.52	-0.75	-0.41	-0.42	-0.75	-0.13
State – Maranhao	0.10	-0.25	0.451	0.63	-1.21	1.91
State – Tocantins	-0.40	-0.58	-0.23	-0.04	-1.34	0.90
State – Pará	-0.11	-0.2	0.02	0.01	-0.53	0.43
State – Roraima	-0.49	-0.6	-0.4	-0.09	-0.64	0.31
State – Amapá	-0.31	-0.4	-0.2	-0.30	-0.96	0.20
State – Acre	-0.10	-0.2	-0.04	-0.09	-0.41	0.17
State – Rondônia	-0.01	-0.1	0.04	-0.08	-0.34	0.12
State – Amazonas	-0.18	-0.2	-0.1	-0.19	-0.43	-0.01
State – MatoGrosso	0.00	-0.04	0.04	-0.1	-0.29	0.04

4.4.2) Model validation

In both the pre- and post-PPCDAM scenarios, the models had apparently strong predictive power with mean AUC values of 0.92 in the first year. In the pre-PPCDAM scenario, which had a longer period of model validation, we found that the predictive power of the calibrated parameter set declined through time to 0.86 over the first 8 years of model predictions (2002 – 2010).

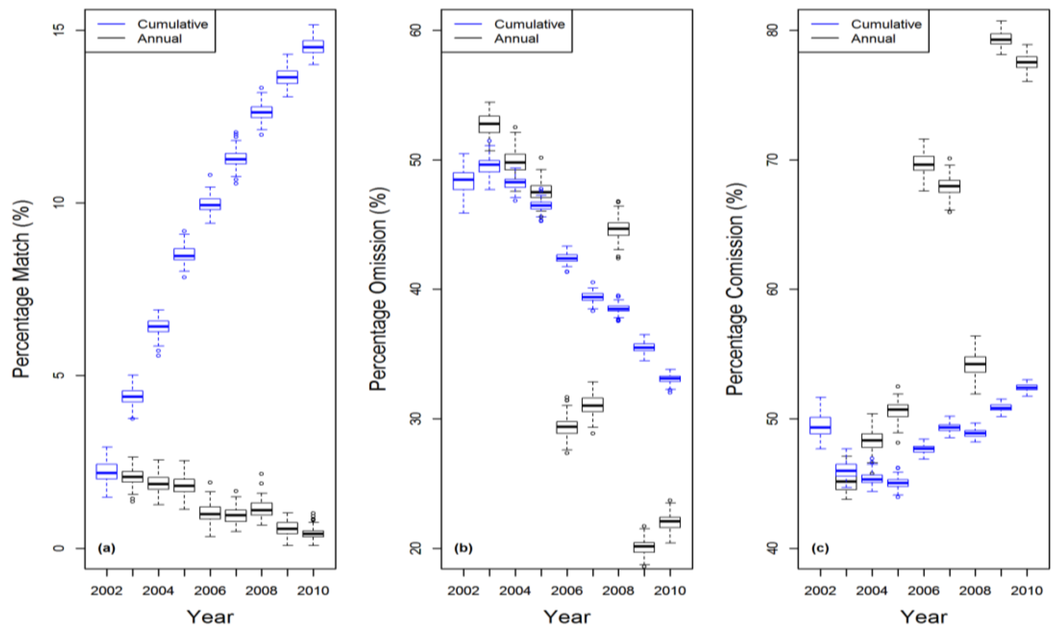


Figure 13 – Pre-PPCDAM model validation results comparing pixel by pixel predicted and observed deforestation between 2002 and 2010. Three validation statistics are presented: (a) mean percent of perfect match; (b) errors of omission; and (c) errors of commission. Validations were conducted in two ways. The 'annual' validations compare predictions from a single year with observations for that same year, whereas the 'cumulative' validations compare all deforestation predictions up to and including that year with observations of cumulative deforestation over the same time period. Variation in these values arises from the 100 model iterations.

In addition, for the pre-PPCDAM scenario, we compared the model predictions with observed data pixel by pixel, both annually and cumulatively from 2002 through 2010. Although most land cover change modellers prefer to compute statistics that compare model outputs with those from a random distribution, such as the Kappa-family of metrics(Soares-Filho et al. 2006; Lapola et al. 2011), we believe that making more demanding pixel-by-pixel comparisons is a more informative and more direct representation of how accurately the model predicts the actual rate and spatial patterns of deforestation(Pontius & Millones 2011). On an annual basis, the model predictions perfectly matched an average of just 2% of observed deforestation events in 2002 and this percentage dropped further through time (Fig. 13a). However, the cumulative prediction accuracy improved through time, for the time span that validation data is available, with an average of 15% perfect match between predicted and observed deforestation by 2010 (Fig. 13a).These results suggest that the model is correctly predicting the general spatial pattern of deforestation, but that the ability to predict the exact sequence of deforestation events is very poor. The majority (>60%) of all annual observed deforestation fell within 25 km of predicted deforestation, and virtually all observed deforestation (84 to 94%) was within 50 km of predicted deforestation. There was no significant change through time in these values (Fig. 14a). When analysed cumulatively (Fig. 14b), an average of 80% of all observed deforestation from 2002 through 2010 occurred within 10 km (2 pixels) of deforestation predicted by the model.

Annual rates of omission errors closely tracked the rate of deforestation (Fig. 13b), with the model omitting more deforestation events in years with more deforestation and omitting fewer deforestation events in years with less deforestation. Commission errors were high and followed the opposite pattern, with most pixels that were predicted to be deforested in a particular year not being observed (Fig. 13c). When validated against cumulative patterns of predicted and observed deforestation commission errors increased through time whereas omission errors decreased, again indicating that the emergent spatial patterns of deforestation are reliable but that the exact sequence in which pixels and deforested is poorly predicted (Fig. 13b and c).

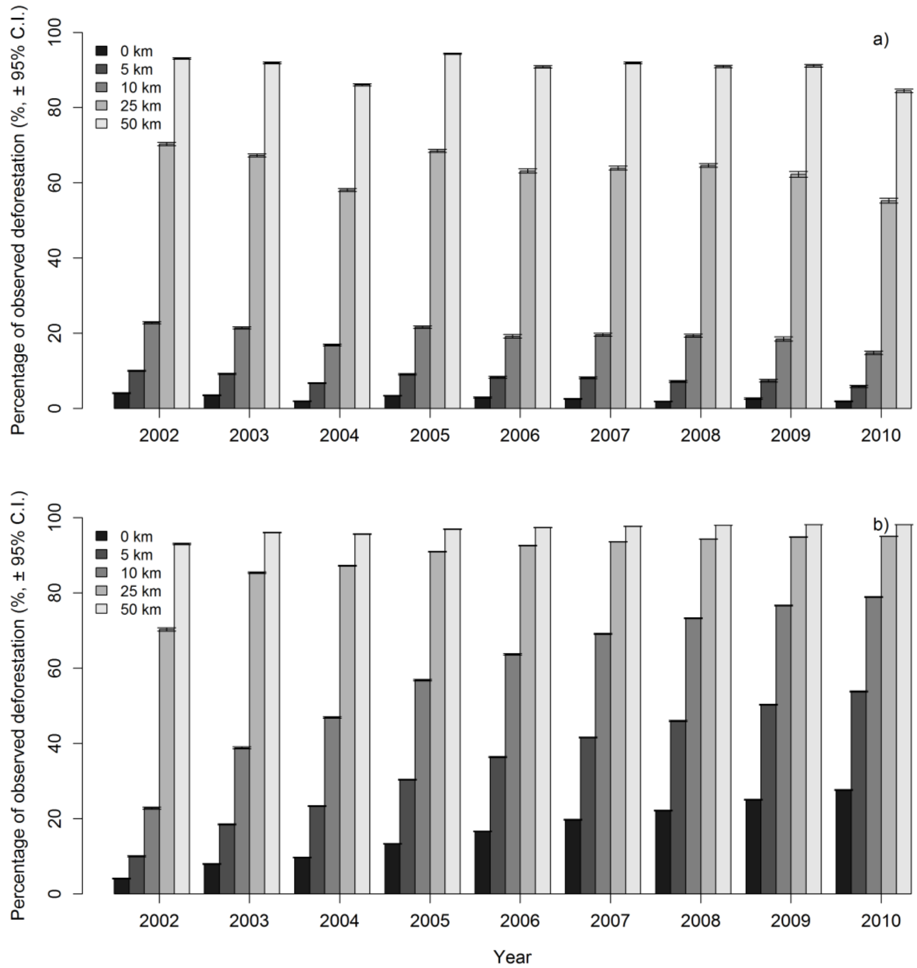


Figure 14 – Pre-PPCDAM model validation showing the spatial dependence of model accuracy. Values represent the proportion of (a) annual and (b) cumulative observed deforestation from 2002 through 2010 that fell within a threshold distance from predicted deforestation.

4.4.3) Rate and location of land-cover change

The total amount (or rate) of deforestation in any given year emerged bottom-up from the accumulation of stochastically determined local deforestation events, and predicted that deforestation rates would almost halve by the year 2050 under the pre-PPCDAM scenario (Fig. 15). Annual differences in deforestation rates among model iterations of the pre-PPCDAM scenario were as much as 0.2%, whereas the post-PPCDAM scenario showed a more stable rate through time (Fig. 15). In 2002, our first year of model predictions for the pre-PPCDAM scenario, the model predicted an average deforestation rate of 0.85%, and for 2010 the pre-PPCDAM predicted a deforestation rate of 0.82% whereas under the post-PPCDAM scenario the average was just 0.2%. Model predictions from the first three years of simulations in both pre- and post- PPCDAM scenarios were in line with observations for this region from INPE (INPE 2012c).These results suggest that the Brazilian government's PPCDAM program, helped by the coincident global economic downturn, seems to have been successful in lowering deforestation rates. Because we only have one dynamic variable in the model (deforestation neighbourhood), the predicted deforestation rate dropped almost constantly through time in the pre-PPCDAM scenario, presumably because all pixels near roads that had the highest deforestation probabilities became deforested leaving behind just pixels with relatively low deforestation probabilities. This pattern is less evident in the post-PPCDAM scenario since the predicted rate of deforestation is much lower. Between the years 2010-2050, the difference in deforestation rates between the pre- and post-PPCDAM scenarios suggests that implementation of PPCDAM will have resulted in an average cumulative reduction in deforestation of 389,884 (± 657, 95 % C.I.) km^2.

Using the annual deforestation probability outputs (Supporting Information S2), we mapped the cumulative deforestation probability predicted for 2050 for both scenarios (Fig. 16a,b), and created two video outputs showing deforestation probability accumulating from 2002 (or 2010 if post-PPCDAM) to 2050 (Supporting Information S3).We found that pixels within a short distance from roads had a very high probability of becoming deforested in the next 40 years, but that protected areas play a vital role of inhibiting the spatial expansion of deforestation.

The “wave”, or temporal sequence, of deforestation across the Brazilian Amazon (Fig. 16c,d) suggests that the sequence of deforestation events follows the deforestation probabilities themselves for both scenarios, with deforestation occurring first along the southern and eastern boundaries of the Amazon before spreading along and out from major highways that penetrate the Basin. However, the magnitude of change is much less intense in the post-PPCDAM scenario. Each iteration of our model represented a different possible future, because we allowed for uncertainty in the model parameters and stochastically determined deforestation events, meaning that in different iterations pixels could be deforested in different years. To capture this uncertainty in our predictions of the wave of deforestation, we mapped out a measure of variance, the inter-quartile range, around our estimates of the year in which each pixel was deforested. The median value of the inter-quartile ranges was 15 years, showing a large amount of uncertainty in the exact timing of deforestation events. In general, for both scenarios, model uncertainty was lowest along the Arc of Deforestation and in areas where nearby roads give immediate access to forest. In the most inaccessible parts of the Brazilian Amazon, the inter-quartile range around our deforestation predictions was as high as 40 years.

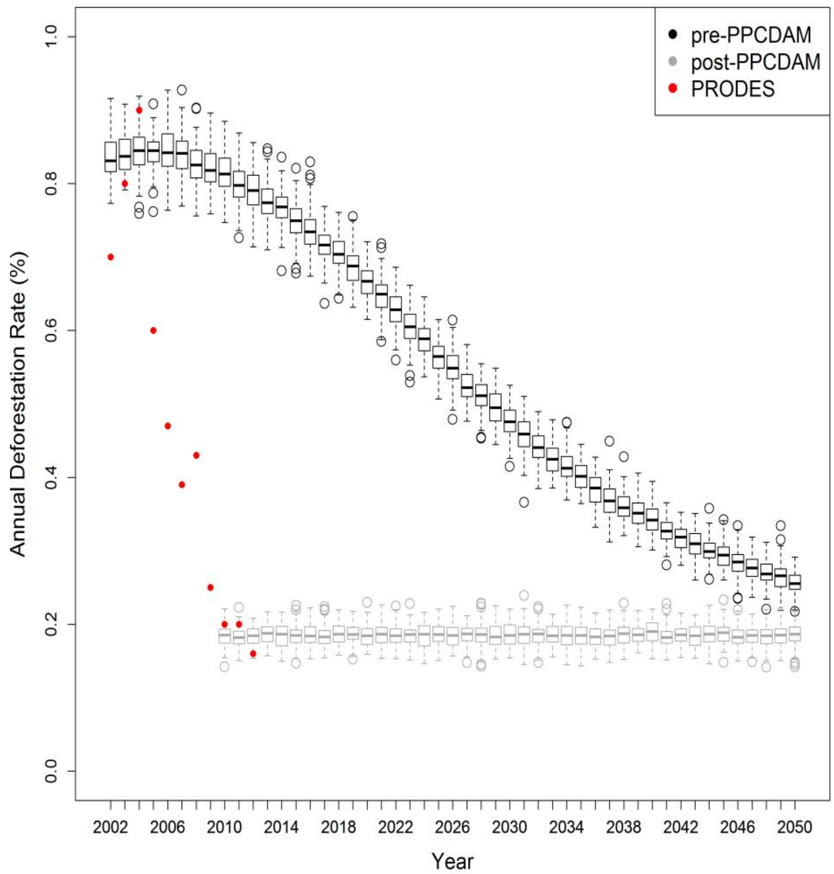


Figure 15 – Predicted deforestation rate in the Brazilian Amazon between 2002 and 2050. Deforestation rates emerged from the local deforestation probabilities in the spatial model, for both the pre- and post-PPCDAM scenarios, and variation in these values arises from the 100 model iterations. Thick lines represent the median, boxes the inter-quantile range and whiskers the maximum and minimum simulated deforestation rates.

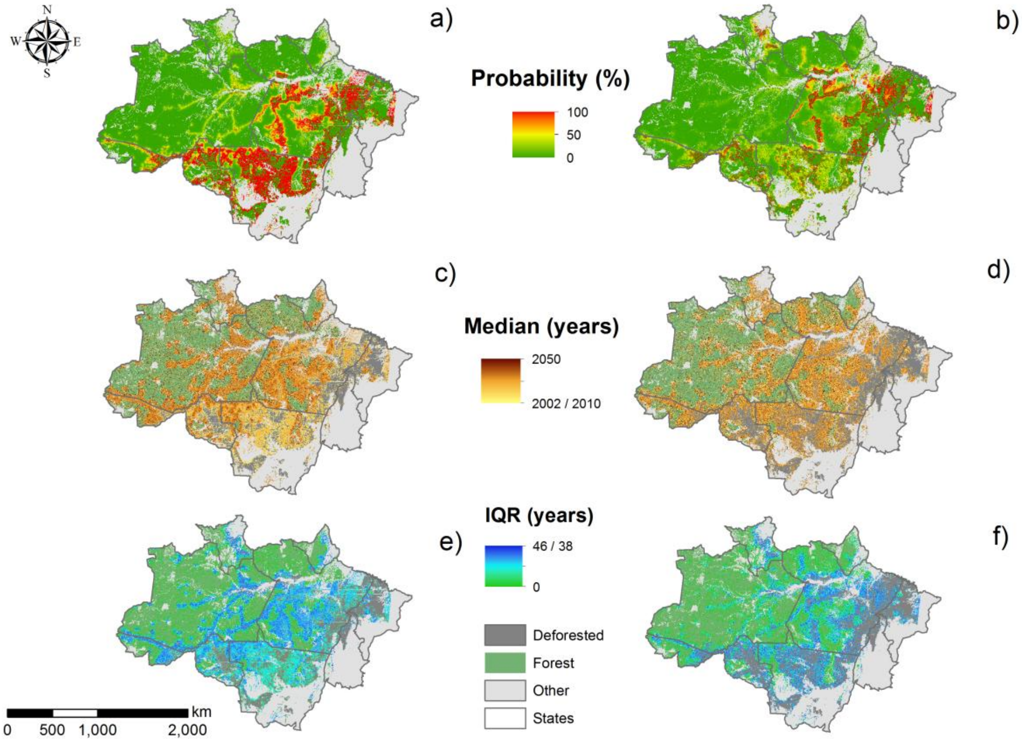


Figure 16 – Deforestation predictions for the Brazilian Amazon under the pre- and post-PPCDAM scenarios. Cumulative deforestation probability in the year 2050 (a) under the pre-PPCDAM and (b) the post-PPCDAM scenario; the wave of deforestation, represented as the median year in which each pixel was deforested (c) under the pre-PPCDAM and (d) the post-PPCDAM scenario; and uncertainty in the model predictions, quantified as the inter-quantile range of the year in which each pixel was deforested (e) under the pre-PPCDAM and (f) the post-PPCDAM scenario. In panels (c,d) and (e,f), measures of central tendency and variation were obtained by comparing model outputs from the 100 model iterations.

4.5) Discussion

With a predicted increase in global human population and, consequently, a rise in external demand for agricultural products, the future of the Brazilian Amazon is at stake if pre-PPCAAM annual deforestation rates prevail in the next decades. In order to predict the potential impacts of deforestation on biodiversity and evaluate the potential effectiveness of conservation strategies, it is vital to be able to accurately predict the magnitude and geographical distributions of future deforestation. However, as we showed the models are not yet accurate enough and it remains vital to improve the prediction of deforestation models. Our model is not the first attempt to make these predictions, but goes beyond previous attempts by capturing three important aspects of deforestation which have poorly been explored in the past: uncertainty, emergence, and contagion. Additionally, we use the stochastic nature of the model to specifically estimate uncertainty around model predictions, which we found to be substantial, by allowing model parameters to vary at each model iteration. Overall our analyses suggest that we can have some confidence in the spatial patterns of cumulative deforestation that will emerge over the coming decades, but that we have little, if any, power to predict the exact sequence of deforestation events at the level of individual pixels. This remains one of the biggest difficulties in land cover change models (Carlson et al. 2012).

Our statistical analysis identified several predictor variables that had demonstrable predictive power for deforestation, but also showed that adding extra predictor variables and parameters to the model does not necessarily lead to a better model. We found that, for both scenarios, the proportion of deforested neighbours has a strong influence on the probability of a given pixel itself being deforested events, which indicates that the "behaviour" of deforestation mimics that of an infectious disease, increasing our confidence in modelling deforestation as a contagious process. Once it starts in a region it can spread very rapidly (Etter et al. 2006a; Boakes et al. 2010) and that spread is even more rapid when roads provide easy access to forests.

Roads play a major role in determining where and how much deforestation will occur and our model has confirmed a large body of literature that emphasises the important impact of roads on deforestation patterns (Geist & Lambin 2002; Fearnside 2008). Regrettably, there is presently a lack of validated models predicting the rate and pattern of expansion of the road network itself in the Amazon (Arima et al. 2008a), making it difficult to include them as a dynamic variable in a deforestation model (Soares-Filho et al. 2006). We can estimate where and when official roads will be created(Soares-Filho et al. 2006), but the same is not true for unofficial roads which are very widespread in the region and represent a major threat to forests

(Brandão & Souza 2006).Given that we allowed the rate of deforestation to emerge from the model itself in a bottom-up manner for both our scenarios, the fact that the predicted deforestation rate drops through time in the pre-PPCDAM is partly, perhaps mostly, an artefact of having a static road network as an input variable in the model, although we know the road network is continuously expanding in this region. We would expect that incorporating a dynamic road network in the model(Arima et al. 2008a), which would continually expand roads through time, would keep deforestation probabilities high and lead to a more steady, or even increasing (as forest would decrease), deforestation rate for both scenarios. However, we also note that economic models of deforestation in the Brazilian Amazon predicted deforestation rates to begin declining under a Business as Usual scenario around 2030 (Soares-Filho et al. 2006), although the reduction predicted there was much lower than predicted by our model. The influence of using static road maps in our model predictions ensures that our predictions of the difference in cumulative deforestation between the pre- and post-PPCDAM scenarios can be considered conservative.

Protected areas strongly constrained the spatial pattern of deforestation in the pre-PPCDAM and post-PPCDAM scenarios, in line with other more direct analyses of the effectiveness of Amazonian protected areas(Nepstad et al. 2006a), and we also found that different types of protected areas exert stronger or weaker limits of deforestation (Nepstad et al. 2006a). Where roads were adjacent to protected areas, we found that deforestation was much more intensive relative to the wider spatial spread of deforestation that occurred around road networks that did not abut protected areas. This suggests, therefore, that the implementation of reserves bounding roads should be supported as an effective means of limiting the spatial spread of deforestation (Soares-Filho et al. 2010).

Within the nine states of the Brazilian Amazon, Mato Grosso, Pará and Rondônia are the three that have had the highest levels of annual deforestation (Aguiar et al. 2007) and this will remain true for the foreseeable future (although there are some early signs of deforestation starting to reach the lower part of the Amazonas state) in both our scenarios, unless governance is improved and there is a strong incentive to restore already degraded lands. According to our pre-PPCDAM simulations, by 2050 Mato Grosso and Rondônia will virtually have no forest left outside the reserves, and the reserves themselves will become very isolated, even though our pre-PPCDAM scenario shows a clear decline of deforestation rates. This pattern is less strong in the post-PPCDAM scenario due to the lower rate of deforestation; however, even here the same states will have the strongest landscape modification. By contrast, even in our most aggressive scenario of deforestation (pre-PPCDAM), Amazonas and Roraima are protected by their inaccessibility, benefitting from the 'passive protection' that arises from their geographic isolation(Rudel 2005; Joppa et al. 2008). However, new planned initiatives to pave roads in all Brazilian Amazon means deforestation will continue to progress, and in particular, in these states(Soares-Filho et al. 2006) the passive protection will be significantly reduced.

Simulation results show that calibrating our model in a different transition year can have a great impact on the rate and location of predicted deforestation. We exploited this variation by calibrating the model for transitions before and after the implementation of PPCDAM, a strong plan to prevent deforestation by the Brazilian government. Our simulations showed a less aggressive scenario of future deforestation both in terms of rate and spatial spread when the model is fitted after the PPCDAM was implemented, when compared to the pre-PPCDAM simulations which were achieved by calibrating the model for a year before the PPCDAM. If we assume that the conditions post-PPCDAM are maintained into the future, we predict this will nearly 390,000 (± 660, 95% C.I.) km^2 of cumulative deforestation by 2050. However, recent changes to the Brazilian Forest Code suggest that the strong level of reduction in deforestation rates between 2004-2010 may not be maintained into the future(Malingreau et al. 2012), although it remains unlikely that rates will climb back to the high values that occurred in the early 2000s.

For both scenarios, because the spatial and temporal patterns of deforestation resulted from stochastic iterations, which also incorporated variation in the model parameters, we were able to capture the temporal and spatial uncertainty in our model predictions. Furthermore, the choice of calibrating our pre-PPCDAM model for the year 2001-2002 enabled us to validate our model predictions for eight consecutive years (2002-2010) both annually and cumulatively, which is rarely done in land-cover change modelling studies. These particular aspects of the model (uncertainty resulting from stochasticity and parameter uncertainty; emergence; contagion; and validation over more than one time step), which have not previously been used together in land-cover change models, allow us to place more confidence on the long term deforestation predictions. The implications from our uncertainty analyses and stringent validations are that we can be more confident in the cumulative pattern of deforestation probabilities through time rather than the exact temporal sequence. However, we are far from accurately predicting the exact sequence in which forests (pixels) will be deforested.

The modelling procedure we have presented here for the Brazilian Amazon under two different scenarios (pre- and post-PPCDAM) can also be used to test the potential impacts of different scenarios, such as the impacts of the construction of new roads and new hydroelectric dams, the implementation of new protected areas, or to estimate biodiversity and carbon losses due to land-cover change. For instance, under the pre-PPCDAM scenario the model predicts rates higher than those observed after the downturn of agriculture in 2005-2006, which is believed to greatly influence the reduction of deforestation in this region [33]. If, however, this economic downturn had translated into less road development the model could dynamically update the road layer and rates would slow down given that the access to forest was stabilizing. Once we re-calibrate the model for the post-PPCDAM scenario using data from 2009-2010, the predicted rates changed considerably and again matched those directly observed by PRODES. However, the largest changes observed in predicted rates between the two scenarios are more directly related to calibration data rather than the emergency property of the model. Therefore, predicting deforestation rates remains the greatest challenge in land cover change modelling. The scenarios presented here were mainly to show the potential of our model structure to quantify the impact of different scenarios, and demonstrate that it can be adapted to address questions about the impacts of policy decisions. Furthermore, tools such as this have potential to be integrated into decision-making processes, providing guidance to conservationists and policy-makers as they plan and test competing land cover decisions. However, these must take into account both the spatial and temporal scales where the model was built and tested, and how the uncertainty in the model output varies at each scale. For instance, models not only can be used to project future trends of deforestation but also to evaluate policy impacts in a long term. However, given that our results showed a low ability to predict the exact temporal sequence of deforestation, we stress the idea that models should provide their users a measure of uncertainty attached to their predictions .We believe that the probabilistic approach we have developed here represents an important step towards the goal of more fully engaging land cover change models with land cover planning decisions.

4.6) Acknowledgments

IMDR is supported by the Grantham Institute for Climate Change and Microsoft Research Cambridge. The work C. Souza Jr. at Imazon has been supported by Gordon and Betty Moore Foundation, the Climate Land Use Alliance (CLUA) and the Skoll Foundation. RME is supported by European Research Council Project number 281986.

Chapter 5 – Evaluating the accuracy of land cover change models: sensitivity to calibration year and temporal changes to deforestation processes

Isabel M. D. Rosa[1], Drew Purves[2], Joao Carreiras[3], and Robert M. Ewers[1]

[1]Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, United Kingdom

[2]Computational Ecology and Environmental Science, Microsoft Research Cambridge, Roger Needham Building, 7 J J Thomson Ave, Cambridge CB3 0FB, United Kingdom

[3]Tropical Research Institute (IICT), Travessa do Conde da Ribeira, 9, 1300-42 Lisbon, Portugal

Manuscript submitted to *Regional Environmental Change* in April 2013

Currently under review.

Keywords: model calibration, Machadinho d'Oeste, land-cover change, predictive models, transition length, temporal trends

Running title: accuracy of LCC models

Corresponding author: Isabel MD Rosa (i.rosa09@imperial.ac.uk)

5.1) Abstract

Land cover change (LCC) models are used in many studies of human impacts on the environment, but knowing how well these models predict observed changes in the landscape is a challenge. Here, we used nearly three decades of LCC maps in Machadinho d'Oeste in the Brazilian Amazon to: (1) investigate how parameter values associated with drivers of LCC vary through time; (2) determine whether the parameter values associated with different land cover transitions (forest to deforested, regeneration to deforested or deforested to regeneration) are more or less variable than each other; and (3) quantify the influence of choosing a particular time period for model calibration on the accuracy of LCC model predictions.

The drivers of LCC varied among transition types, demonstrating broadly that deforestation tends to occur along roads and outside protected areas whereas regeneration tends to occur far from roads and inside protected areas. Our temporal analysis on model parameters revealed some degree of variation among time transitions (e.g. increasing effectiveness of protected areas), but for the majority of parameters there was no significant trend through time. However, the accuracy of LCC model predictions was heavily dependent on the year calibrated, suggesting that a widespread reliance on single calibration time period in the LCC modelling literature may be providing biased predictions of future LCC.

The next generation of LCC models will need to embed trends in parameter values within simulations to allow the processes determining LCC to change through time and exert their influence on model predictions.

5.2) Introduction

Human induced land cover change (LCC) in the tropics is severely altering the landscape and causing, among other aspects, the depletion of many species' habitat (Gibson et al. 2011), increasing the amount of carbon released to the atmosphere (Baccini et al. 2012), and interfering with indigenous peoples otherwise undisturbed (Laurance 1999). LCC models are used in many studies such as those that study the impact of building a new road in a forested area (Soares-Filho et al. 2004) or those that aim to estimate carbon losses due to changes in the land cover (Galford et al. 2010). There are many models that predict future LCC (Verburg et al. 2002 ; Soares-Filho et al. 2006), but knowing how well these models predict the observed changes in the landscape is still a challenge. Here, using a case study in the Brazilian Amazon we tackle this problem by investigating how well these models work and how reliant that is on the timing and time-scale of data.

The Brazilian Amazon is world's largest remaining tropical rainforest but has been under great pressure due to human-related activities for the past few decades. The modern era of Amazonian deforestation began in the 1960s and 1970s with colonisation schemes implemented by the Brazilian government, which aimed to relocate unemployed people from other parts of Brazil to the Amazon. The land grants associated with this process, along with government-sponsored road development that greatly facilitated access to forests, kick-started deforestation in the region (Fearnside 2005). Later, with the healthy Brazilian economy growing fast and the increasing demand for agriculture products from other parts of the world, there was rapid expansion of large-scale agriculture, mainly for the production of beef and soybeans (Nepstad et al. 2006b). In more recent years, however, the Brazilian government policies against illegal deforestation (Soares-Filho et al. 2010), market-based campaigns (Rudorff et al. 2011) and the coincident global economic crisis have reduced the rate of deforestation (INPE 2012c) and altered the spatial pattern of forest clearings (Rosa et al. 2012).

Across the Brazilian Amazon there is considerable variation in the processes underlying deforestation, the rates of deforestation and the emergent spatial patterns of deforestation. Here, we focus on the municipality of Machadinho d'Oeste, covering an area of approximately 8,500 km^2 in the state of Rondônia. This municipality was initially a settlement project implemented by the Brazilian government in 1982 as part of the national development program called Northwest Region Integrated Development Program (POLONORDESTE), however, rapid development and population growth lead to its autonomy in 1988 (Miranda 2012). The majority of people in the municipality are dependent on agriculture for subsistence and less than half live in urban areas (Miranda 2012). With such an intense relationship between people and agriculture it is not surprising that this municipality has been undergoing severe deforestation since the settlement was created in 1982. By 1997 it had lost 30% of its original forest extent (Mangabeira et al. 1998), and by 2005, 23 years after establishment, the original landscape had been transformed into a mosaic of remnant forest patches, secondary vegetation, pastures, agriculture lands and small urban areas (Gomes et al. 2009). The average area of agricultural plots is 46 ha and the main sources of income for families in the region are livestock and coffee (Miranda et al. 2008; Gomes et al. 2009). In line with the rest of Rondônia, the livestock numbers in Machadinho d'Oeste rose sharply from 4 000 in 1989 to more than 215 000 by 2007 (IBGE 2006). Since then the rate of forest loss has declined, and in places even reversed with land abandonment having led to regenerating secondary forest. By 2011, 33% of the municipality was deforested, just slightly higher than the amount registered in 1997 (INPE 2012a).

LCC models are useful tools that can be used to provide future simulations of landscape modification under specific scenarios. These models examine and statistically define the spatial patterns of LCC in a particular time period, and use these observations to make future predictions by extrapolating them into the future. The choice of time period is, in most instances, limited by which data is available (*e.g.* satellite images in a particular year are covered with clouds making it hard to classify), while sometimes it can be tied to the desire of studying the effect of a particular climate phenomena such as El Nino (Ramos da Silva et al. 2008), the effect of implementing new protected areas (Soares-Filho et al. 2006) or paving a road (Soares-Filho et al. 2004), among others. Since these models are heavily dependent on the input data used to calibrate the model, one of the main limitations of LCC models is that the parameters used in its future simulations are essentially frozen in time, with the statistical description of LCC patterns estimated at a single time period and implicitly assumed to remain constant into the future. The fact that the parameters are not updated through the simulations means that the effect of a variable, such as distance to roads or the effectiveness of protected areas, for instance, would remain the same through time. This is a problem because we do not know how this artefact of freezing time and assuming that the processes that drive LCC do not change propagates through model predictions to generate errors in LCC predictions.

Here, we take advantage of a large historical LCC database encompassing nearly three decades of LCC in Machadinho d'Oeste to quantify the degree to which freezing parameters in time can influence the accuracy of predictions arising from LCC models. Specifically, the objectives of our study were to: (1) investigate how parameter values change through time and thereby determine the constancy of the statistical patterns of LCC that are fed into predictive models; (2) determine whether the parameter values associated with different land cover transitions (forest to deforested, regeneration to deforested or deforested to regeneration) are more or less constant than each other; and (3) quantify the influence of choosing a particular time period on the accuracy of LCC predictions. Together, our analyses are designed to examine the robustness of the modelling techniques currently used to predict LCC, with a view towards developing models that appropriately quantify the uncertainty in LCC predictions.

5.3) Materials and Methods

The land cover dataset of Machadinho d'Oeste mapped forest cover at 21 points in time during the period 1984 to 2011 (Carreiras et al. 2012). Using a high frequency time series of Landsat 5 Thematic Mapper (TM), they classified 30 m spatial resolution Landsat 5 Thematic Mapper (TM) satellite images into three classes: mature forest (hereafter referred to as forest), secondary succession forest (regeneration) and non-forest (deforested). To model the probability of LCC in each year we used five explanatory variables, which include the most important proximate causes of deforestation in the region: location of previous deforestation (contagion) (Alves 2002), distance to roads (Pfaff et al. 2007), distance to rivers (Pfaff 1999), distance to settlements (Pfaff 1999) and protected areas (Nepstad et al. 2006a). The roads (both official and unofficial) and protected areas datasets were obtained from the Instituto do

Homem e Meio Ambiente da Amazônia (Imazon), and the rivers and settlements maps were obtained from the Instituto Brasileiro de Geografia e Estatística (IBGE). For areas of regeneration, we also included three metrics generated from the time series itself: the period of active land use (PALU) prior to land abandonment, the age of regenerating forest (ARF) and the frequency of clearance (FC), which is the number of times a pixel of forest/regeneration was cleared (deforested) until a particular year. Other variables that are commonly used in Amazonian LCC models such as dry season length, soil fertility and altitude were not included in this analysis because the spatial scale of the municipality is too small to find any significant differences that could influence the spatial patterns of LCC. All data were converted to 30 m cell size to match the land cover dataset, using a Universal Transverse Mercator (UTM) coordinates system (zone 20 S), WGS-84 datum.

Our first analysis was to investigate how the set of parameters that described the impact of explanatory variables on LCC (*e.g.* distance to roads, the effect of protected areas) vary among LCC transition types and through time. We calculated the proportion of the landscape occupied by the three land cover classes at each of the 21 time points for which we had land cover maps, and determined the rate of change for each transition type in each of the 20 time periods separating adjacent time steps (1984-1986, 1986-1987, 1987-1989, 1989-1990, 1990-1991, 1991-1994, 1994-1995, 1995-1996, 1996-1997, 1997-1998, 1998-1999, 1999-2001, 2001-2003, 2003-2005, 2005-2006, 2006-2007, 2007-2008, 2008-2009, 2009-2010, 2010-2011). We used linear regression models to test whether these changes were significant through time.

The data collected and described above were separated into two categories of variables: static (roads, rivers, settlements and protected areas) and dynamic (land cover and proportion of deforested/regeneration neighbours) variables (Soares-Filho et al. 2002). Static variables represented features that are assumed to stay constant through time (*e.g.* rivers) or that we lack information to be able to updated them through time (*e.g.* roads) and were only calculated once in the beginning of the modelling process. Dynamic variables, by contrast, represent features that change through time and were re-calculated at the beginning of each model time simulation.

Our LCC modelling approach is based on that of Rosa et al. (under review-b), who developed a dynamic and spatially-explicit model of deforestation to predict the potential magnitude and spatial pattern of deforestation. It differs from previous models in three ways: (1) it is probabilistic rather than deterministic, allowing quantification of uncertainty around the predictions; (2) the rate of LCC emerges 'bottom up', as the sum of local-scale deforestation probabilities driven by local processes, as opposed to being imposed 'top down'; and (3) LCC is modelled as a contagious process, such that local rates of LCC increase through time if adjacent locations have experienced recent LCC (Figure 17 illustrates the modelling procedure).

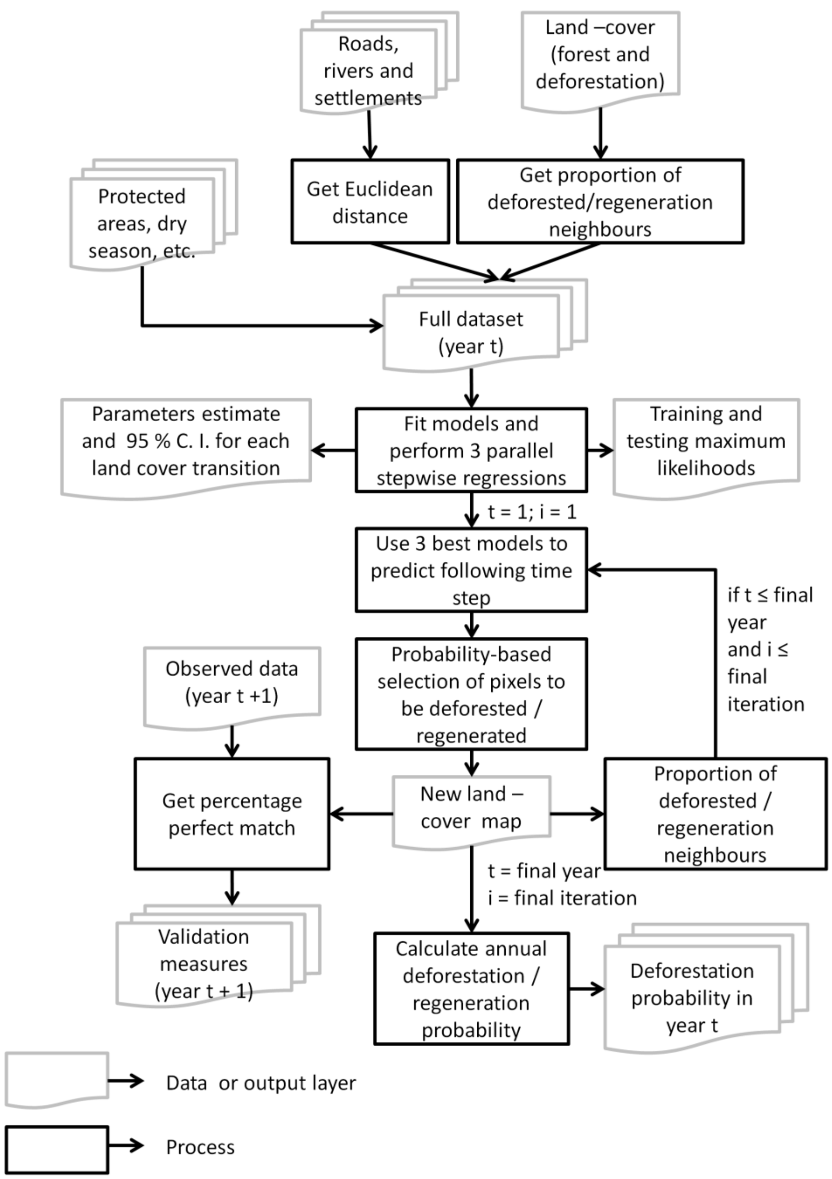


Figure 17 – Modelling procedure flowchart. The flowchart illustrates the construction and running of the land cover change model. *i* refers to the model iteration, and *t* is the year being modelled.

The fact that the dataset had three land cover classes allowed us to model three specific LCC transitions rather than the single transition of forest to deforested implemented by Rosa et al. (under review-b). Here, we constructed independent models of forest to deforested (FtoD), regeneration to deforested (RtoD) and deforested to regeneration (DtoR). Given a land cover map of time *t* with three land cover classes (forest, regeneration and deforested), for each pixel of forest at time *t*, the model calculates the probability of being deforested at time *t+1*; for each pixel of regeneration the model calculates the probability of being deforested at time *t+1*; and finally, for each deforested pixel the model calculates the probability of becoming regeneration at time *t+1*.

First, we fitted all single driver models to explain the effect of each predictor variable on each LCC transition in 19 of the possible time periods, ignoring the first time period (1984-1986) because the 1984 landscape included just two of the land cover types (forest and deforested). Using as an example the transition FtoD, the model is based around $P_{defor,x,t}$, the probability that cell *x* becomes deforested (or is converted to regeneration in the case of the DtoR transition - $P_{reg\ x,t}$) in a set interval of time *t*. This probability is defined as a logistic function:

$$P_{defor,x,t} = 1/(1 + \exp[-\kappa_{x,t}]) \qquad (1)$$

such that as $\kappa_{x,t}$ goes from minus infinity to plus infinity, $P_{defor,x,t}$ goes from 0 to 1. Then we wrote simple linear models for $\kappa_{x,t}$ as a function of the driver variables affecting location *x* at time *t*. In this analysis the driver is only one of the predictors at a time. Once the parameter value was found we used linear regression models to test for any trend that would indicate the parameter values had changed through time or in relation to the rate of LCC.

Having a large land cover dataset allowed us to test if using different calibration years (the year of the initial maps used for model calibration), different transition lengths (the number of years between the initial and final map used in model calibration), validation year (the year used for model validation), and the number of time steps the model is extrapolated into the future (number of time steps until year used for validation) have important impacts on the ability of LCC models to predict future land cover. To do this, we used all prior land cover maps to parameterise the set of 40 models (out of a total of 66 possible models) that generated predictions for the year 2011 (Fig. 18). For example, the six-year transition period 1991-1997 models land cover in six-yearly intervals giving predictions for 1997, 2003 and 2010, but not 2011, and therefore could not be used in our analyses. By contrast, there are three time periods beginning in 1991 that do predict land cover in 2011 and were used: 1991-1995 (2011 will be the 5th time step predicted), 1991-1996 (4th time step predicted) and 1991-2001 (2nd time step predicted).

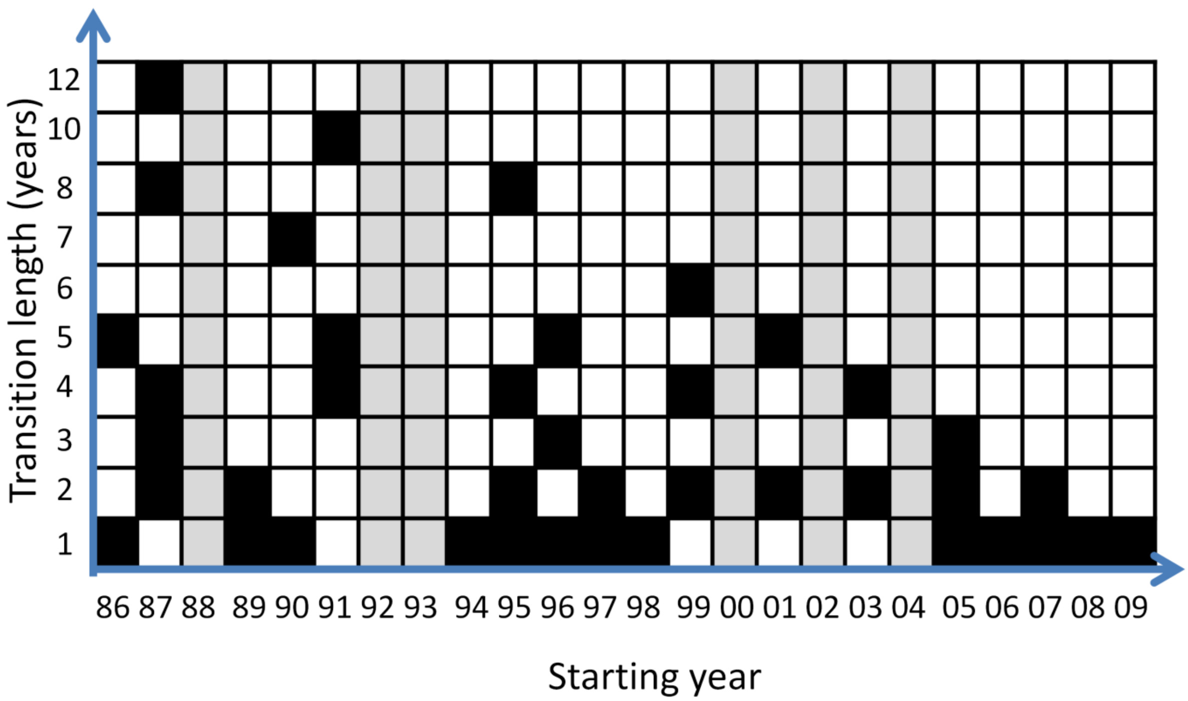


Figure 18 – Illustration to show the combination of land cover maps and transition lengths that allow the model to predict of a land cover map for 2011 (in black). Grey bars show years with no data available.

For each of the 40 models, the process started with a forward stepwise regression where at each step, and for each land cover transition, models differed only in the combination of static and dynamic variables included in the definition of $\kappa_{x,t}$ (Rosa et al. under review-b), which is the linear model function of the driver variables affecting location *x* at time *t*. In this analysis all drivers are considered at the same time. We used the library 'Filzbach' (http://research.microsoft.com/en-us/projects/filzbach/) to return, for each parameter, a posterior probability distribution using Markov Chain Monte Carlo sampling techniques. From these distributions we extracted the posterior mean, and a credible interval, given the model structure and the data. These intervals will later be used to draw parameter values for each model iteration, allowing the quantification of uncertainty around our predictions. Afterwards, to carry out the parameter estimation, all that was necessary was to define the log-likelihood:

$$\ell(X|s,\theta) = \sum_{x,t} log\{Z_{x,t}P_{defor,x,t} + (1 - Z_{x,t})(1 - P_{defor,x,t})\} \quad (2)$$

where $Z_{x,t}$ is the observed deforestation at location *x* at time *t*, and *s* refers to one of the models considered. At each step of the forward stepwise regression, a cross-validation was carried out by parameterising the model against a randomly selected subset of 50 % of locations (50 % of forest pixels when the transition being modelling is FtoD, 50 % of regeneration pixels when the transition being modelling is RtoD or 50 % of deforested pixels when the transition being modelling is DtoR) and calculating the training likelihood, then calculating the test likelihood on the remaining 50 % of the locations (forest, regeneration or deforested, depending on the transition being modelled), using eq. 2, for validation.

The best models were selected as the ones with the maximum test likelihood for each of the 120 stepwise regressions (40 models times three land cover transitions). These best models for each land cover transition were then used to run 40 separate simulations of LCC, each of which predicted the spatial pattern of land cover until 2011. In order to do so, we re-applied eq. 1 in each time step, recalculating the dynamics variables (*e.g.* fractional deforestation/regeneration adjacent to each location *x*), and using a slightly different set of parameter values at each iteration (to incorporate parameter uncertainty) drawn using a Gaussian distribution from the credible interval determined above. This provided an updated $P_{defor,x,t}$ (or $P_{reg\ x,t,}$ - probability of regeneration – depending on the land cover transition being modelled) for each location *x*, which was then deforested (or converted to regeneration) with that probability. In practise, this was implemented as follows: for each *x*, draw a random number from a uniform distribution bounded at 0 and 1, deforest (or convert to regeneration) *x* if this number is less than $P_{defor,x,t}$ (or $P_{reg\ x,t}$). After these deforestation/regeneration events were implemented, $P_{defor,x,t}$ and $P_{reg\ x,t}$ was calculated for every location *x* again, allowing for another round of land cover change. Given that the Rosa et al. (under review-b) modelling procedure has a very strong stochastic component with each individual deforestation/regeneration event is drawn randomly using a weighted probability, each simulation was repeated for 100 iterations which allowed us to construct confidence intervals around our model predictions. For each simulation, we output the predicted probability per land cover transition, the annual change on each of the land cover transitions and the new land cover map.

Finally, to assess how well each of these 40 LCC models was able to predict the actual pattern of land cover we employed a pixel-by-pixel validation metric called perfect match, which tests if the model was able to predict the exact location of a land cover on that specific year. We chose this metric because, although it can be considered more rigid when compared to neighbourhood metrics such as the Kappa-family (Pontius & Millones 2011), it is more informative in showing how well the model is predicting change. The perfect match metric compares the predicted and observed maps on a pixel-by-pixel basis (30 m cell size). First, using 2011 as an example of validation year, the observed change in land cover between the two observed maps (initial land cover and 2011) was calculated: if *x* is 1 (forest) in the initial map and 0 (deforested) in 2011, gets the value 1; or if it is 2 (regeneration) in the initial map and 0 in 2011 gets the value 2; and finally, if it is 0 in the initial map and 2 in 2011 gets the value 3. Then, we calculated the predicted change by following exactly the same procedure using the predicted maps of 2011 (the 100 land cover outputs from the 100 iterations of the model) instead of the observed map. Finally, we compared the 100 predicted maps of change against the observed map of LCC. When the value of *x* (1, 2 or 3) is the same in both the observed and predicted change maps, the pixel gets the value 1, otherwise it gets the value 0. In the end, we summed all pixels with value 1 and divided this number by the total amount of observed change, including all land cover transitions. This methodology avoids validating pixels that do not change during the modelled period, which would inevitably lead to high but unrealistic values of perfect match. Combining the results of the 100 iterations we calculated the mean percentage of perfect match across iterations. Every land cover map available was used as validation year, which means that each model was validated against several land cover maps in time. For instance, the 2005-2006 model was validated for five years (2007, 2008, 2009, 2010 and 2011); the only exception is the last model (2009-2010) that could only be validated for a single year (2011).

5.4) Results

The landscape in Machadinho d'Oeste was highly dynamic during the period 1984-2011, with large temporal changes in the proportion of the municipality occupied by each of the three land cover classes (Figure 19a). There was a strong decrease in the proportion of area occupied by forest (slope = -2.43, t = -36.22, R^2 = 0.99, df = 19, p<0.001) that was accompanied by an expected increase in the proportion of area occupied by the other two classes (deforestation: slope = 1.48, t =17.6, R^2 = 0.94, df = 19, p<0.001; regeneration: slope = 0.93, t = 7.91, R^2 = 0.78, df = 19, p<0.001). However, there was high annual variability in the rates of each LCC transition type (Figure 19b), with no significant temporal trends detected for any transition type (|t| < 1.7, R^2 < 0.14, df = 17, p >0.09).

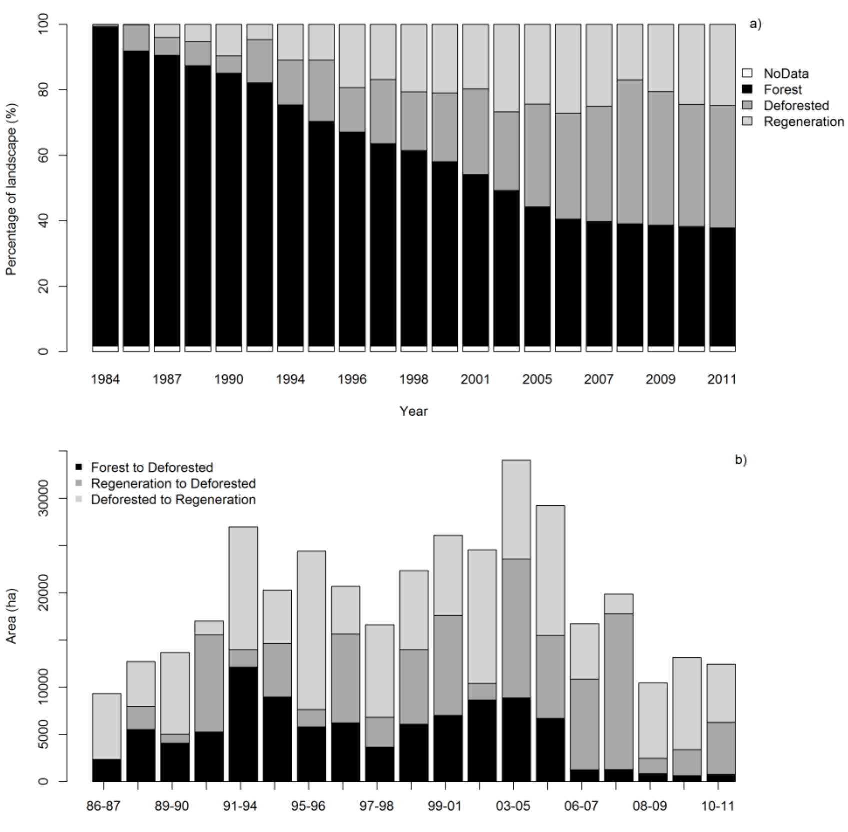


Figure 19 – (a) Proportion of forest, regeneration and deforestation between 1984 and 2011 in Machadinho d'Oeste; NoData refers to a common water mask that was applied to all the dates in the time-series and (b) rate of change in each of the three land cover transitions: forest to deforested (FtoD), regeneration to deforested (RtoD) and deforested to regeneration (DtoR).

Parameter values found in the single driver analysis supported expected trends, with estimates varying among the three transition types (Fig. 20). Deforestation happens closer to roads and settlements but afforestation happens further away from human presence (Fig. 20c and e, respectively). Protected areas play an important role inhibiting deforestation and favouring regeneration (Fig. 20f). Furthermore, all three transitions presented some degree of spatial autocorrelation (Fig. 20b) given by the positive value of the parameter associated with contagion, although this pattern was found to be stronger in the FtoD transition.

Few parameters exhibited temporal trends, suggesting their effects were relatively constant through time (Fig. 20). Exceptions were protected areas' ability to prevent deforestation, which gets stronger through time (slope = -0.04, t = -7.27, R^2 = 0.76, df = 17, $p < 0.001$) suggesting that while the forest area shrinks the effectiveness of protected areas tend to become stronger; the contagious effect of deforestation in the RtoD transition showed a strong positive trend through time (slope = 0.08, t = 4.54, R^2 = 0.55, df = 17, $p < 0.001$) as well as the spatial autocorrelation of regeneration in the DtoR transition (slope = 0.21, t = 7.79, R^2 = 0.78, df = 17, $p < 0.001$). Finally, the effect of the frequency of clearance also revealed a significant positive trend through time in the transition RtoD (slope = 0.02, t = 3.48, R^2 = 0.21, df = 17, $p = 0.003$). This indicates that deforestation from regeneration tends to happen in areas that have been more frequently cleared in the past. Most parameter estimates were robust to the rate of change, meaning that years with higher amounts of land changing did not result in big changes in the parameter values. The only exception was for the transition DtoR in relation to distance to settlements, where in years with higher amounts of regeneration tended to happen further away from human-related activities (slope = 2.82×10^{-10}, t = 4.82, R^2 = 0.58, df = 17, $p < 0.001$).

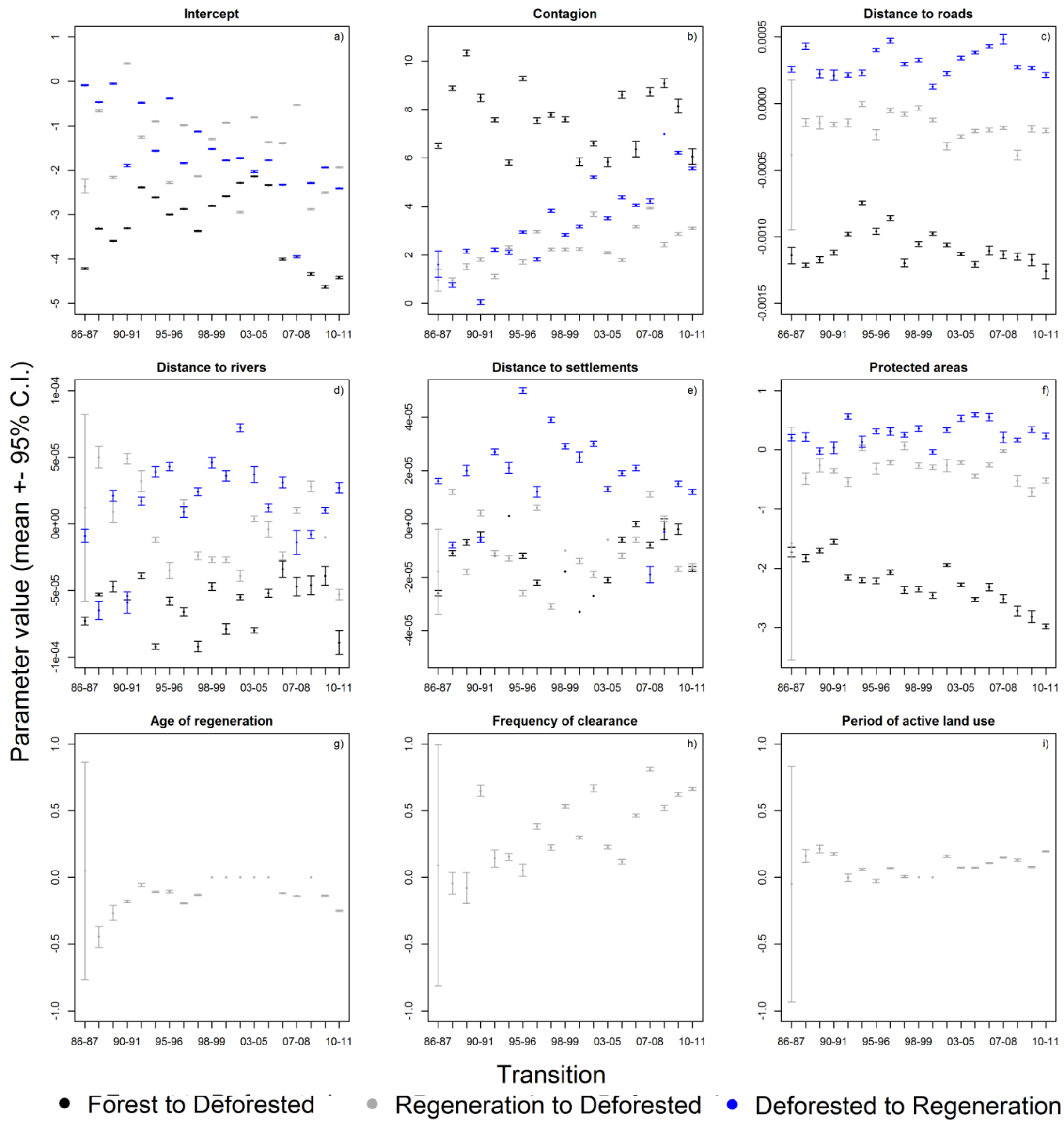


Figure 20 – Parameters' mean values (plus 95% confidence interval) for all one-parameter models fitted between 1986 and 2011. The year 1984 was omitted from this analysis because it only includes two land covers (forest and deforestation), not allowing the three land-cover transitions. In the 1986 map, the area of regeneration is very small which leads to higher uncertainties (larger confidence intervals) in the parameters values found for this transition in this year.

Parameter estimates from the 120 stepwise-based models were consistent with the results from the single driver models (Table 4). Contagion, protected areas and distance to roads were consistently important, but their effects on DtoR differ from the other two transitions. In accordance with previous results from the single driver analysis, this indicated that land abandonment (or regeneration) tends to occur away from roads and inside protected areas, whereas deforestation is very closely linked to the presence of roads and protected areas are effective in preventing its occurrence. Rivers were sometimes important but with no consistent direction of effect, while settlements rarely mattered except for DtoR transition, where its inclusion represented that land abandonment tends to occur further away from settlements. Finally, regarding the three landscape metrics only present in the models of the transition RtoD, ARF was found to be the most regularly present on the best models, followed by FC. Younger areas of regeneration and areas that are frequently cleared are the ones more likely to get deforested again.

Table 4 – Results from the 40 modelling procedures: number of times each parameter was included in the best model (given in % of the 40 models), used for simulation on each land cover transition (forest to deforested (FtoD),regeneration to deforested (RtoD), deforested to regeneration (DtoR)); out of these, the number of times the parameter's mean was found to be positive (given in % of the number before); average and standard error of the parameters' values in the best models for each land cover transition.

	Land cover transition	Intercept	Contagion	Distance to roads	Distance to rivers	Distance to settlement	Age of regeneration	Frequency of clearance	Period of active land use	Protected areas
Included in best model (%)	FtoD	100	100	95	45	2.5	x	x	x	100
	RtoD	100	100	90	60	0	60	47.5	27.5	42.5
	DtoR	100	77.5	77.5	37.5	37.5	x	x	x	37.5
Positive percentage (%)	FtoD	5	100	0	16.67	0	x	x	x	0
	RtoD	27.5	92.5	0	54	0	0	89	64	0
	DtoR	17.5	55	100	80	93	x	x	x	87
Average	FtoD	-2.00	4.00	-0.00056	-0.00003	-0.00001	x	x	x	-1.48
	RtoD	-1.00	1.16	-0.00016	0.00001	0	-0.21	0.21	0.05	-0.36
	DtoR	-1.42	-0.65	0.00035	0.00002	0.00004	x	x	x	0.15
Standard error	FtoD	0.19	0.34	0.000023	0.0000043	0	x	x	x	0.07
	RtoD	0.23	0.17	0.000013	0.0000076	0	0.03	0.03	0.02	0.04
	DtoR	0.21	0.42	0.000018	0.0000067	0.0000047	x	x	x	0.06

Figure 21 and Table 5 show the results obtained from the model used to test whether the calibration starting year, the number of years between calibration years (transition length), the year being validated and the time step of the model being validated had an effect of how well the model was able to perfectly predict the observed land cover map. We found that model accuracy was significantly affected by the initial year and validation year, whereas the length of the transition period and, surprisingly, the number of time steps in the future had no significant effects (Table 5). As expected, these results suggest that the choice of initial year to calibrate will have a strong impact on the ability of the model to predict a particular landscape. Model accuracy improved with later starting years as well as later validation years (Fig. 21a and d). Intuitively, this indicates that when trying to predict a land cover pattern it is better to use maps from a date closer to the one we are trying to predict. Furthermore, we found that contrarily to what would be intuitive, using a larger transition length does not necessarily translate into more accurate predictions (P-value > 0.05, Fig. 21b).

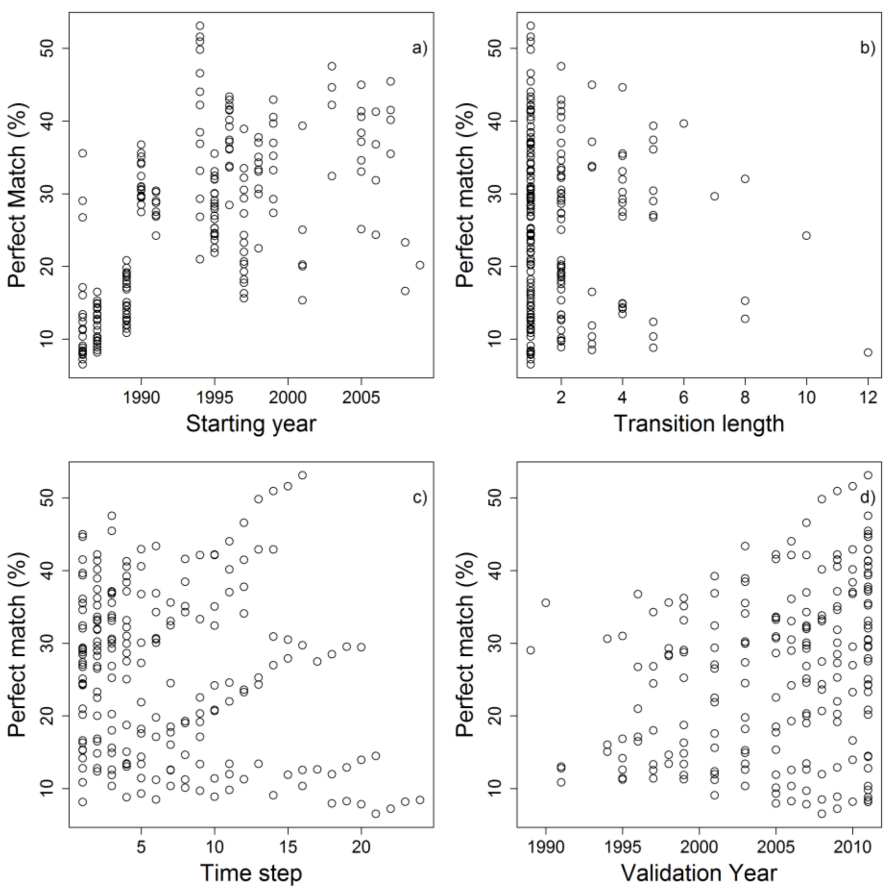


Figure 21 – Perfect match (%) results obtained with the 40 modelling procedures by comparing land-cover change predicted and observed between model's year of calibration and validation year (total number of validations made equals 270, because a single model can be validated in several years).

Table 5 – Statistical analysis of predictions perfect match results for the 40 modelling procedures, when compared to observed land cover in the year being validated, as a function of initial calibration year (Initial year), the length of the time transition being used to calibrate the model (Transition length), the model time step being validated (Time step) and finally the year being validated (Validation year).

Variable	**Degrees of freedom**	**Sum of squares**	**Mean Sum of Squares**	**F-value**	**P-value**
Initial Year	1	10 270	10 270	185.166	< 0.001
Transition length	1	7	7	0.119	0.731
Time step	1	118	188	2.121	0.147
Validation year	1	223	223	4.018	0.046
Initial Year x Transition length	1	657	657	11.840	< 0.001
Initial Year x Time step	1	4 280	4 280	77.179	< 0.001
Initial Year x Validation year	1	15	15	0.264	0.608
Transition length x Validation year	1	1	1	0.022	0.882
Time step x Validation year	1	4	4	0.072	0.788
Initial year x Transition length x Time Step	1	224	224	4.036	0.046
Initial year x Transition length x Validation year	1	3	3	0.048	0.828
Initial year x Time step x Validation year	1	168	168	3.022	0.084
Transition length x Time step x Validation year	1	158	158	2.842	0.093
Initial year x Transition length x Time step x Validation year	1	14	14	0.255	0.614
Residuals	192	10 649	55		

5.5) Discussion

The landscape in Machadinho d’Oeste has changed dramatically during the last 30 years. The amount of primary forest dropped by almost one third between 1984 and 2011. However, this decline in the area of primary forest has slowed down since the early 2000s, which is probably due to a lack of good quality wood in the remaining primary forest patches that has led to a reduction in logging activity (Miranda et al. 1999). In addition, the remaining primary forest in 2011 is almost entirely sited within indigenous lands where logging and any other land use commonly found outside these areas such as croplands is forbidden. Deforestation is still an ongoing process in the region, but the forest that is being cleared is more typically regeneration, or areas that have been deforested at least once in the past. This suggests that the main deforestation agents in the region are now farmers rather than loggers, and that the balance between regeneration and deforested land is maintained by farmers temporarily abandoning pastures and allowing regeneration to establish, before re-clearing the land in a cycle designed to maintain soil fertility in these nutrient-poor landscapes (Smith et al. 1999). The positive value found for the parameter associated with the frequency of clearance (Fig. 20h) and the fact that this value gets higher through time add some support to this result by suggesting that areas that have been cleared once in the past are more likely to get cleared again.

The analyses of the three LCC transition types can be interpreted as examining the actions of different deforestation agents: loggers are better represented by the transition FtoD, while farmers are better represented by the transition RtoD. The last transition (DtoR) can mainly be assumed to represent the process of land abandonment by farmers since no reforestation program is implemented in the region. The differences among these processes were reflected in the different parameter values determined for the three transition types. In both transitions that lead to deforestation, forest and regeneration areas next to already deforested areas were confirmed to be more likely to be deforested in the near future (Fig. 20b), with spatial contagion being one of the most important factors in determining the rate and location of LCC. In the transition that predicts regeneration, however, this pattern was less evident suggesting that the process of regeneration does not follow a contagious process as strongly as deforestation does. The role of roads in Amazon deforestation has been highlighted many times in past studies (Nepstad et al. 2001; Brandão & Souza 2006; Pfaff et al. 2007) and our study found similar results. What is interesting to note is the marked difference between land cover transitions (Fig. 20c), with the road effect being much stronger for the deforestation of primary rather than regeneration forest, and land abandonment tending to occur far from roads. Usually, logging companies provide the first access to forests by opening roads (Arima et al. 2005a), as such deforestation tend to happen closer to these roads, and farmers take advantage of these roads but their decisions about where to deforest are also mediated by other factors such as soil fertility (Escada et al. 2005). Protected areas have a strong preventive effect on deforestation in the Brazilian Amazon (Nepstad et al. 2006a), and in Machadinho d'Oeste we found that this effect strengthened through time as the remnant forest was progressively restricted to protected areas (Figure 20f). Protected areas have only a limited role in preventing the deforestation of regeneration, largely because regeneration is mainly located outside of protected areas. Land abandonment (DtoR) tended to happen more inside protected areas than outside, which suggests that no real effort is being made to let the deforested area recover to a state closer to the original forest outside these protected areas. As such, although forest cover seems to be improving inside protected areas, these might become very isolated and their future can be undermined (Ribeiro et al. 2006).

The temporal analysis on the model's parameters revealed some degree of variation within time transitions, but for the majority of the parameters no significant trend through time. This lack of variability can partially be explained by the fact that some of these variables simply do not change through time (such as the rivers and settlements) as it does not change their relation with the land cover process being modelled. In addition, due to lack of data, we were unable to update the road map annually, meaning that we treat one of the most important and more dynamic drivers of deforestation as static. This most likely affects the results found for the parameter associated with this variable. The exceptions were the protected areas, which became increasingly more important in preventing deforestation of primary forests, while the landscape changes and the forest outside protected areas gets heavily degraded; the contagion effect of deforestation (RtoD transition) and regeneration (DtoR transition) which become increasingly more important showing how the land cover change strongly follows a contagion pattern; and the frequency of clearance, which shows that while there was a lot of primary forest unprotected, people tend to deforest areas of primary forest, but as the landscape evolves and these forests become scarce, leaving almost only primary forests inside protected areas, people opt to re-deforest areas that have already been cleared at least once in the past and are now regeneration.

The accuracy of LCC models was heavily dependent on the year in which models were calibrated (Table 5), suggesting that a widespread reliance on single calibration time periods in the LCC modelling literature (Michalski et al. 2008; Maeda et al. 2011; Yanai et al. 2012) may be providing biased predictions of future LCC. The reason for this dependence of model accuracy on calibration year is likely to be that the processes underlying LCC change through time. We found that models calibrated in the 1980s, soon after the first establishment of settlements in Machadinho d'Oeste, were almost invariably poor at predicting LCC, whereas those calibrated from the mid-1990s and later were generally much better. This suggests that accurate predictions of future landscapes cannot be made using maps from early in the past as model inputs, because the processes determining LCC are dynamic and change through time. As such, parameter included in the model that predict future landscape should allow variation which would reflect this changes rather than being kept constant assuming that the same process of LCC perpetuates through time.

In the case study presented here, LCC started with large areas of primary forest being lost and converted to non-forest areas. Some of these areas recovered to areas of secondary vegetation which were then frequently cleared, while others were permanently converted to agriculture and urban areas. Then, this LCC process started to slow down and actually 'stabilised' at the beginning of 2000s which explains why the starting year was found to be only important until mid-1990s, the period of time when there was higher rates of LCC. This same trend in landscape alteration can also explain why the validation year was found to be significant and, especially, why validations done on later years were found to be the ones resulting in higher model accuracy. Given the contagious behaviour of LCC, small changes in an already heavily degraded landscape are proportionally easier to capture by the model than capturing high/low amount of change in a pristine landscape, where there is more possibilities of change to happen.

The calibration year is the most important factor influencing the accuracy of the LCC models, however, once this is controlled the time step and the transition length become important as well in explaining the accuracy of the model. This suggests that after choosing the most recent land cover maps for modelling future landscape alterations, we achieved higher values of perfect match if we validate our annual predictions against land cover maps that are from a date closer to the one modelled (shorter time step). The validation results tend to get worse (or more uncertain) if we validate our predictions against a map from a distant date in time (larger time step). Finally, a similar result was found for the transition length, meaning that after choosing the most recent maps of land cover for modelling, a shorter transition length lead to better model's performance rather than a longer transition length. This result can be counterintuitive as people usually think that a larger transition length might be best for model calibration because a short transition length might not be representative of the dynamics of land cover in the region, rather it might represent a spontaneous event. However, the pattern was found to be less strong than the one found for the time step and, as such, more work needs to be done in order to further analyse this relationship.

The main objective of most LCC models is to provide realistic simulations of future LCC that can help policy-makers to make decisions on questions such as where to best place a new protected area, or to assess the impact of building a new road. In order to do that, models need to be able to replicate the different land cover changes that are observed on the ground and simulate the role of the different agents responsible for those changes, and provide information about its degree of prediction variability. This process is greatly complicated if the role of different agents changes through time, and our analyses suggest that these temporal changes are both real and strongly influence the predictive accuracy of LCC models. The implication is that this temporal variability in parameter estimates needs to be incorporated into models, although this represents a considerable challenge to a modelling discipline that rarely quantifies the uncertainty around their predictions (Rosa et al. under review-a; Rosa et al. under review-b). For parameters that exhibit non-directional variability, this might be possible by running stochastic models that sample parameter values from the statistical distributions of parameter estimates rather than using the 'best' estimate alone. Parameters that exhibit directional trends, however, will require more extensive time series data to allow that trend to be quantified. The next generation of LCC models will need to embed those trends within simulations in order to allow the processes determining LCC to change through time and exert their influence in the model's predictions. To do this, first the parameters' trends should be indentified from the available LCC dataset and then these trends could be sampled in each model run to produce the land cover predictions with an associate measure of error to each parameter *ergo* each LCC process.

5.6) Acknowledgements

IMDR is supported by the Grantham Institute for Climate Change and Microsoft Research Cambridge; JMBC by Foundation for Science and Technology (FCT, Portugal) REGROWTH-BR Project (Ref. PTDC/AGR-CFL/114908/2009). The authors thank Imazon for providing the data on road networks for the Brazilian Amazon.

Chapter 6 – The carbon legacy of modern tropical deforestation

Isabel M. D. Rosa[1], Matthew Smith[2], Drew Purves[2], and Robert M. Ewers[1]

[1]Imperial College of London, Silwood Park Campus, Buckhurst Road, Ascot SL5 7PY, United Kingdom

[2]Computational Ecology and Environmental Science, Microsoft Research Cambridge, Roger Needham Building, 7 JJ Thomson Ave, Cambridge CB3 0FB, United Kingdom

Manuscript in preparation to be submitted to *Nature/Science* in 2013

Keywords: tropical deforestation, spatial probabilistic model, land-cover change, plant carbon, carbon legacy, committed emissions

Running title: carbon legacy of tropical deforestation

Corresponding author: Isabel MD Rosa (i.rosa09@imperial.ac.uk)

6.1) Abstract

Tropical deforestation is responsible for a significant share of the land-cover change (LCC) global carbon emissions. In developing countries these emissions are usually higher than those from fossil fuel burning, however, in the global carbon budget, estimates from LCC have the lowest precision associated. Particularly, uncertainties include the actual forest extent (and how it changes through time) and the spatial distribution of biomass.

In order to calculate emissions from current deforestation it is necessary to take into account the history of the landscape. Combining a probabilistic model of deforestation, a plant carbon model, and a widely used bookkeeping carbon (C) model, we estimate carbon emissions due to gross deforestation from 1950 to 2009 for the Amazon, the Congo Basin and Southeast Asia. Region-specific models were developed, explicitly including uncertainties in both the forest extent predictions and in the spatial distribution of plant carbon used to determine the amount of carbon lost.

Here we show how deforestation started in mainland Southeast Asia, but quickly the Amazon became responsible for the highest share of emissions. Further, we show that even if deforestation had ended in 2010 there were still large quantities of carbon to be released ('the carbon legacy'). From Amazonian deforestation there would be 2.70 Pg C (± 0.70, 95% C.I.) to be released, whereas from the Congo basin there would be 0.82 Pg C (± 0.15, 95% C.I.) and 0.34 Pg C (± 0.08, 95% C.I.) from Southeast Asia. The amount of carbon released immediately is higher than the one committed for future release in the first few years of analysis, but by the end of the time period (2009) these accounted in every region for at least two-thirds of the total carbon emissions.

6.2) Introduction

Tropical forests, despite covering less than 10% of world's terrestrial area, are the largest pool of species richness and biodiversity (Myers 1988), and play an irreplaceable role in the global carbon cycle (Malhi & Grace 2000). In addition, they provide a variety of fundamental ecosystem services such as water (Vourlitis et al. 2002) and climate regulation (Laurance 2004), and habitat for unique animals and plants (Myers et al. 2000). The Amazon is the largest remaining continuous tropical forest and accounts for 49% of total tropical forest carbon stock, followed by Southeast (SE) Asia which accounts for the 26% and the Congo Basin which accounts for the remaining 25% (Saatchi et al. 2011). The future of tropical forests is, however, a topic of great concern, due to the extensive environmental damage already caused by increasing food production (Foley et al. 2005). During the last decades they have been facing severe human pressure, especially due to the rapid expansion of large-scale agriculture (Gibbs et al. 2010), but also due to commercial logging and associated infrastructure developments, such as road building (Laporte et al. 2007).

Landscape dynamic in the tropics is already impacting the global atmosphere, with deforestation being responsible for 10-20% of global carbon dioxide (CO_2) emissions (van der Werf et al. 2009). In many tropical nations, land use change emissions have a higher share in the nation's carbon budget than the energy/industry (Houghton & Hackler 1999; Minnen et al. 2009). Estimates of tropical forest losses and resultant carbon emissions found in the literature are, however, highly variable. In the global carbon budget, carbon emission estimates from land-use change have the lowest precision associated (Malhi et al. 2002). This is due to the uncertainties in the true deforestation rates (Grainger 2008), the amount of biomass and soil carbon held within different forest types (Keith et al. 2009), as well as the spatial distribution of forests (Achard et al. 2004), and the methodologies employed to convert deforestation into carbon emissions. This uncertainty in the amount of forest cleared annually, together with variable biomass estimates, leads to considerable uncertainties in estimating the amount of carbon lost due to land cover changes (Holly et al. 2007).

The first source of uncertainty concerns the highly variable rates of deforestation in the tropics (Geist & Lambin 2002; Ewers et al. 2008). In the Amazon, fluctuations in the rates of deforestation are related to several factors such as the economic health of the countries, infrastructure development, and the world's demand for agriculture products, such as beef or, more recently, soybeans (Morton et al. 2006; Nepstad et al. 2006b; Ewers et al. 2008). The Amazon and tropical SE Asia share a similar deforestation history. In the late 1960s/early 1970s, government-sponsored colonization schemes resulted in the opening of new roads that allowed the access of a large number of people to tropical forests otherwise undisturbed (Geist & Lambin 2002; Fearnside 2005; Rudel 2005). Prior to the 1990s, rural populations increased greatly with the resettlement programs, who establish themselves along the roads opened by logging companies (to extract timber) (Rudel 2005). After 1990, the pattern changed, with agribusinesses becoming much more important in explaining deforestation rates and leaving the forest heavily fragmented (Asner et al. 2009; Rudel et al. 2009). Low labour costs, high commodity prices and vast resources of land lead to rapid expansion of soy plantations in the Amazon and oil palm plantations in SE Asia, with both crops occupying previously forested land.

The Congo Basin has been spared the same extent of large-scale deforestation observed in the Amazon and SE Asia, mainly because of its inaccessibility (Rudel 2005; Rudel et al. 2009). Prior to the 1990s, deforestation was mainly driven by logging and small-scale subsistence farming, however, it is hard to understand the extent of the colonisation impacts in the central African forests due to lack of systematic data (Wilkie & Laporte 2001).This is the most challenging region to monitor due to the persistence of cloud cover and the fine-scale of forest cover change dynamics (Duveiller et al. 2008). Nonetheless, it is believed that this is the tropical region with the lowest deforestation rates. In recent decades human-related activities have steadily increase their impact on the land-use dynamics (Rudel et al. 2009) and industrial logging became the most extensive form of land use in this region (Laporte et al. 2007).

The second source of uncertainty concerns the estimates of carbon biomass in tropical forests (Houghton 2003). Rather than using an average biomass for a region it is important to consider the spatial distribution of biomass, since carbon emissions from deforestation are obviously determined by the biomass present in the forests that are actually deforested (Houghton 2005). Recently, two groups developed biomass maps (Saatchi et al. 2011; Baccini et al. 2012) which address this issue and helped making great progress in estimating land cover change emissions; these have since been used to update estimates on carbon emissions (Baccini et al. 2012; Harris et al. 2012a). However, since these maps are built using current biomass datasets, areas that were deforested in the past will necessarily show low biomass values, which might have been high before they were deforested. This is a limitation when accounting for historical emissions.

A third source of uncertainties in estimates concerns the different methodologies applied when calculating carbon emissions (Houghton et al. 2012). Some authors opt to simply multiply the forest loss by the carbon content (Soares-Filho et al. 2006; Harris et al. 2012a) whereas others attempt to incorporate time-lags by differentiating between immediate loss and loss by the decay of remaining biomass (Houghton et al. 2000; Baccini et al. 2012). In a review of 13 estimates of net land use emissions that cover the period, authors found estimates to vary within a range of 0.5 up to almost 2.5 $PgC.yr^{-1}$ during the period we will cover in our analysis (1950-2009). Globally, emissions from land use / land cover change in the 1980s and 1990s varied greatly, averaging 1.14 and 1.13 $PgC.yr^{-1}$ (± 0.23 s.d.), respectively (Houghton et al. 2012). For the 2000s, two groups of researchers converged on a lower estimate of 0.81 $PgC.yr^{-1}$ (Baccini et al. 2012; Harris et al. 2012a; Harris et al. 2012b).

One important aspect of calculating carbon emissions from deforestation that is often missed is to take into account the history of the landscape (Ramankutty et al. 2007). A deforestation event that happens today leads to a future release of carbon either from the decay of forest products, or from the decay of slash left at the site in the subsequent years (Houghton et al. 2000). Therefore, estimating the amount of carbon that has accumulated since modern deforestation started is vital to reduce the uncertainty regarding global carbon emission due to deforestation in the current days. Unfortunately, past deforestation emissions are hard to estimate due to lack of observed data on land cover change before the 'era' of earth observation through satellite begun in the 1970s (Boyd and Danson, 2005).

Here we used a stochastic spatially-explicit model (Rosa et al. under review-b) to estimate gross deforestation rates from 1950 through 2009 in the three main tropical regions of the world: the Amazon, the Congo Basin and SE Asia. As data for 1950s is unavailable we start the model using observed data from the 2001-2010 time period and back-cast our model predictions until 1950, only allowing one land cover transition (deforested to forest) in the model. Then, using these annual predictions of forest change, we combined a well-known bookkeeping carbon model (Houghton 2005) with a spatially explicit model of potential plant carbon (Smith et al. 2012) to estimate resultant carbon emissions. In the end, we estimated what we term 'the carbon legacy' of tropical deforestation: the amount of carbon left to be released due to past deforestation, even if all human-related activity in the tropics causing deforestation were to immediately halt.

6.3) Data and Methods

We collected all the necessary data to include in the spatial model from a variety of freely available sources (Appendix B – Data sources). These included annual global land cover maps from 2001 through 2010 derived from MODIS satellite imagery (https://lpdaac.usgs.gov/products/modis_products_table/mcd12q1, from which we extracted the forest and deforested extent and annual rates), a map of altitude, a dry season length (derived from a precipitation layer) map, roads and rivers networks maps (from which we calculated the distance to nearest feature) and a population density map. The cell size used in this study was 0.01^{o} (~ 1 km^2) and the spatial reference was WGS-84 datum.

We used the deforestation modelling approach developed by Rosa et al. (under review-b) to 'reconstruct' deforestation history from 1950 through 2009 in the three main tropical regions: Amazon, Central Africa and SE Asia. This spatial model has three main features: (1) it is probabilistic rather than deterministic, allowing quantification of uncertainty around the predictions and parameters associated with drivers of change; (2) the rate of deforestation emerges 'bottom up', as the sum of local-scale deforestation probabilities driven by local processes, rather than being imposed 'top-down'; and (3) deforestation is modelled as a contagious process, such that local rates of deforestation increase through time if nearby locations that have experienced recent change. Due to characteristic (3) we opted to 'divide' the region of SE Asia into six sub-regions that are continuous and allow deforestation to behave in a contagious matter which would be harder if a single model for the whole region was implemented. In total, eight models were developed: one for the Amazon, one for Central Africa and six for SE Asia (one for Mainland SE Asia, one for Borneo, one for Sumatra, one for Papua, one for Sulawesi and one for the archipelago of Philippines).

Usually, models are used to predict the probability of future deforestation ($P_{defor\ x,t}$) in *t+1* within a forested area (forested pixels) given a set of drivers of change (e.g. distance to roads) in *t*. Here, however, we 'invert' this and determine the probability of a deforested area in *t* having been forested ($P_{for\ x,t}$) in *t-1*. The model is then based around $P_{for\ x,t}$, the probability that a deforested cell *x* becomes forest in a set interval of time *t* (Appendix B – Modelling past deforestation). Further, having nine years of land cover dataset allowed us to avoid relying on a single time transition to make our predictions and thus incorporate uncertainties in the rates of change by combining simulations from all possible transitions in order to backcast deforestation from 2010 through 1950.

The data were divided into two types of variables: static variables (*e.g.* altitude, rivers), those that are assumed to remain constant through time or that we know are dynamic but for which no datasets are available with multiple updates through time (*e.g.* roads); and dynamic variables (forest cover map and proportion of forested neighbours), those that we assume change through time. We generated a total of 72 models (9 time transitions x 8 regions/sub-regions), each starting with a forward stepwise regression where at each step, models differed only in the combination of static and dynamic variables included in the model (Rosa *et al.*, under review). Using the C++ library 'Filzbach' (http://research.microsoft.com/en-us/projects/filzbach/), a posterior probability distribution using Markov Chain Monte Carlo sampling techniques, was returned for each parameter. From these distributions we extracted the posterior mean, and a credible interval, given the model structure and the data. These intervals were later used to randomly draw parameter values for each model iteration, allowing the quantification of uncertainty around our predictions (Appendix B – Modelling past deforestation).

The best models for each time transition and each region were selected as the ones with the maximum test likelihood. These best models on each region and time transition were then used to run separate simulations of 'forestation', each of which predicted the spatial pattern of forest extent backwards in time until 1950. Given that the Rosa et al. (under review-b) modelling procedure has a very strong stochastic component with each individual 'forestation' event being drawn randomly using a weighted probability, each simulation was repeated for 100 iterations which allowed us to construct confidence intervals around our model predictions, and reduce the reliance of our predictions on the parameter estimates gained from observing a single transition time period. The annual 'forestation' probability map in each time step (1950-2009) was calculated as the number of times a pixel was selected to be converted to forest in any given year, divided by the total number of iterations (N = 100). For each simulation, we output the predicted probability of 'forestation', the annual change in forest extent and the new land cover map.

To validate our model predictions, we employed a stringent pixel-by-pixel validation method and determined the perfect match, omission and commission, both annually and cumulatively for each region. Further, to account for near/far misses (Pontius Jr et al. 2004b) we calculated a distance-based metric of validation, again, annually and cumulatively (Appendix B – Validation of model predictions 2001-2009). Further, we compared the annual rates of deforestation that our model predicted with published estimates found in the literature for each region.

In order to calculate carbon emissions due to deforestation from1950 through 2009 in each of the three regions, we needed to combine our forest extent predictions with a map of biomass per region. However, most biomass maps are built using current data (Saatchi et al. 2011; Baccini et al. 2012) and as a result the areas that are currently deforested have a much lower value of carbon content than the one that would have been observed in 1950. To overcome this issue we used the terrestrial carbon model developed by Smith et al. (2012), which outputs a median (and confidence interval) map of potential plant carbon. These maps do not treat deforested areas as areas with low biomass values; instead it incorporates climate dependencies and other biophysical parameters to calculate the potential carbon content in any area of the globe. Furthermore, like before, we explicitly include uncertainty in the carbon maps by drawing a set of 100 maps, one to be used in each model iteration, from the median and uncertainty maps produced by Smith et al. (2012). This was done by using a log-normal distribution with specific input parameters (mean and standard deviation) for each pixel. The histograms presenting the distribution of plant carbon in each of the three regions are shown in Figure B1 (Appendix B), and were found to be lower than other authors' estimates (Baccini et al. 2012).

Carbon emissions were calculated by overlapping these plant carbon maps with our deforestation predictions from 1950 through 2009, for each model iteration (N = 100). Then, using a widely used bookkeeping carbon model (Houghton 2005; Le Quere et al. 2009; Baccini et al. 2012), once a pixel of forest is cleared, carbon was lost (1) immediately (20%) and (2) through time by the decay of slash left in the site (70%, at a decay rate of 0.1/yr), resultant forest products (8%, at a decay rate of 0.1/yr) and elemental carbon (2%, at a decay rate of 0.001/yr). Regrowth and soil carbon losses were not included in the analyses, the former because the level of data is insufficient to perform the same type of analysis that is done from primary forests, and the latter because soil carbon response have different trajectories and it is thought to be small when compared to the losses caused by the removal of aboveground biomass (Harris et al. 2012a). At each model year and iteration a record of the amount of carbon lost immediately and by decay was kept, allowing us to separate carbon emissions into those that were caused by deforestation in that year (immediate emissions) and those that were caused by deforestation before that year (committed emissions, or the carbon legacy of past deforestation).

6.4) Results

6.4.1) Parameter values found for the models estimating past deforestation

In future deforestation predictive models we expect to find a negative parameter associated with distance to deforestation and the presence of protected areas. Here, as we 'inverted' this type of models for back-casting, then a positive value is expected, meaning that forests will 'grow' further away from, for example, roads until eventually they reach the roads where it all started. Indeed, that was what we observed for most models across regions and time transitions (Tables B1 to B8). In some cases (see for example Table B3 in Appendix B), the parameter associated with roads was negative, usually this happened in calibration years closer to 2000. This suggests that in these calibration years most observed deforested area was located close to roads leaving no pixels available to 'grow' back to forest further from these roads. This is also true for the Congo basin (Table B2) where, presently, the amount of deforested area is much smaller when compared to the other two regions – Amazon and SE Asia – and thus, still more concentrated along roads.

One major difference observed was that neither population density, altitude nor dry season length appeared to play an important role in determining the conversion of land cover in the Amazon (Table B1 in Appendix B). This means that once the model incorporated the effect of roads and proportion of forested neighbours (which mimics the expansion of croplands), the other variables available to be chosen no longer improved the predictive power of the model and, therefore, were not chosen for the final model. This is probably due to the high density of roads found in this region when compared to Congo basin (Table B2) and SE Asia (Tables B3 to B8). In addition, due to the contagion aspect of the land cover change model and the geography of the region, we opted for modelling forest loss in SE Asia at a sub-regional scale using six models. This means that the sub-regions being modelled in SE Asia as well as the Congo basin are always much smaller than the Amazon, which might have influenced the choice of number of variables included in the final model.

The parameter associated with the contagiousness of land conversion (proportion of forested neighbours) was found to be the only parameter that for all models was always chosen to determine the location and rate of change (Tables B1 to B8 in Appendix B). Next, parameters associated with accessibility were very commonly chosen, especially the distance to roads, which is in line with what we know about the drivers of tropical deforestation (Brandão & Souza 2006; Laporte et al. 2007; Gaveau et al. 2009b). Further, the existence of protected areas, which help prevent deforestation, was very frequently selected which is also in line with the land cover change modelling literature in these regions (Nepstad et al. 2006a; Duveiller et al. 2008; Gaveau et al. 2009a). Interestingly, parameters varied substantially between regions but there was very little variation through time within a region.

6.4.2) Predicted deforestation rates

In accordance to the history of modern tropical deforestation that can be found in the literature, model estimates show a rapid increase in deforestation rates (km^2/yr) from 1950 through 2009 (Fig. 22) especially in the Amazon since the 1970s (Rudel 2007) and SE Asia since the 1990s (Zhao et al. 2006). Deforestation in the Congo basin started later and at a slower pace (Fig. 22a and b). In proportional terms, the fastest rate of deforestation was in SE Asia, followed by the Amazon and finally the Congo basin. Further, we note that the model predicts a reduction in deforestation rates after 2004 in the Amazon, which is coincident with the period of time when deforestation rates reduced significantly in the Brazilian Amazon, which represents 60% of the region (Nepstad et al. 2009). In addition, we found that in the calibration years (2001-2002 through 2009-2010), SE Asia was the region with the highest percentage of area already deforested (32% ± 0.31 95% C.I.), followed by the Amazon (18% ± 0.55 95% C.I.) and finally the Congo basin (11% ± 0.49 95% C.I.).

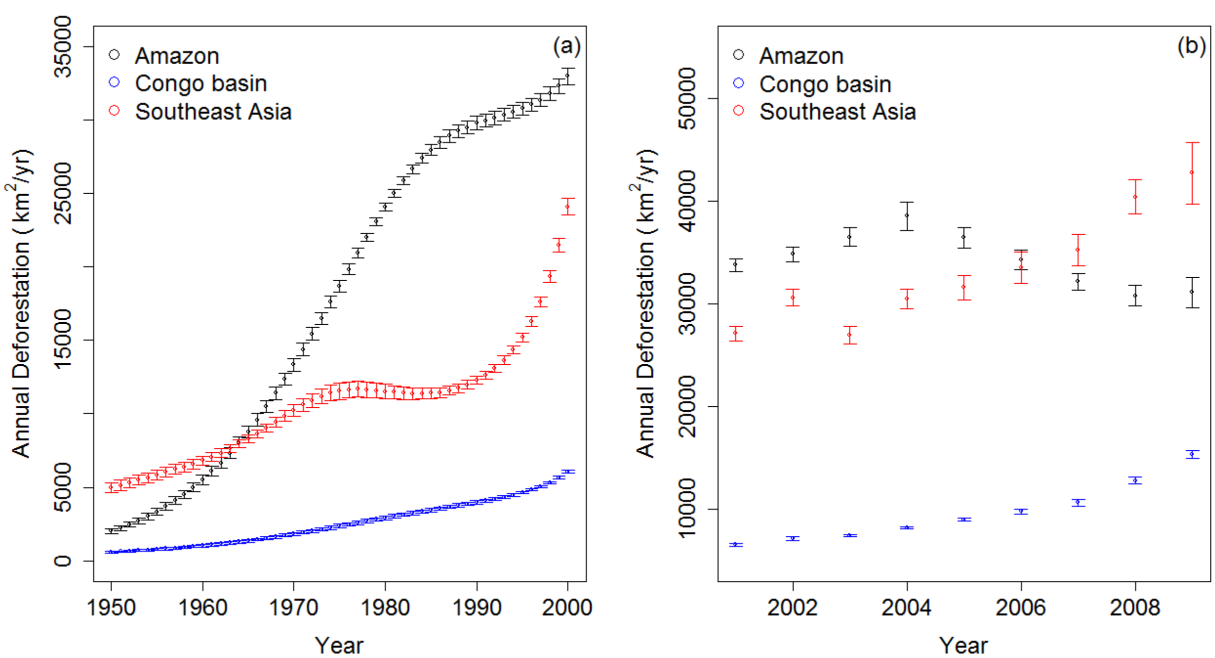


Figure 22 – Predicted annual deforestation rate (km^2/yr) from (a) 1950-1951 through 2000-2001 – model simulation years – and from (b) 2001-2002 through 2009-2010 – model parameterisation years. The 95% confidence interval was calculated using the results from the 100 iterations, for each model for each sub-region.

6.4.3) Past emissions from tropical deforestation

Annual carbon emissions from gross tropical deforestation rose sharply from virtually no emissions in 1950 to nearly 1.0 Pg C/yr in 2009 (Fig. 23a). Estimate for present day (2000s) were found to average 0.82 Pg C/yr (± 0.04, 95% C.I.) which is in line with the consensus achieved recently by two other research groups (Harris et al. 2012b) for 2000-2005. The Amazon is, on average, responsible for nearly 50% of these emissions (Fig. 23b), followed by SE Asia (35%, Fig. 23d) and the Congo Basin forests (15%, Fig. 23c).

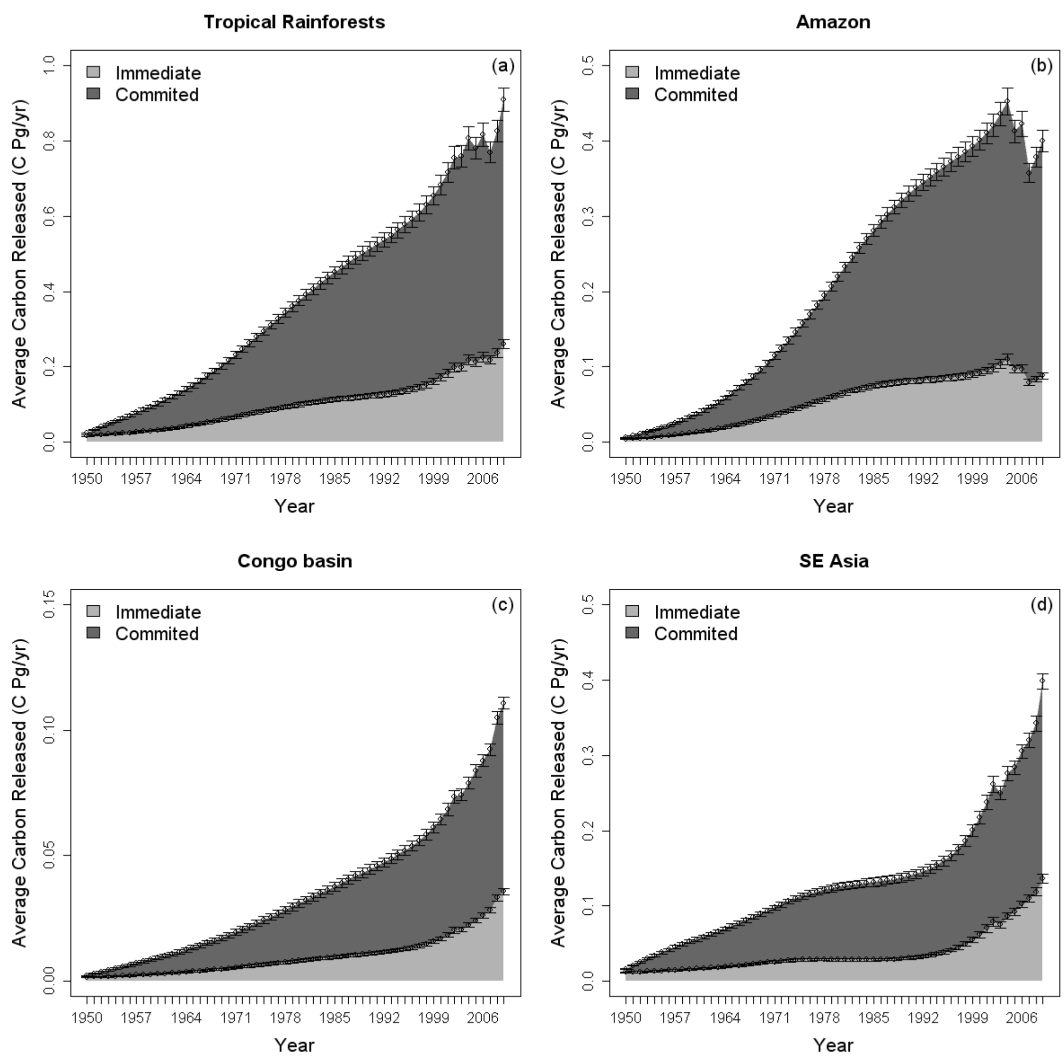


Figure 23 – Carbon lost immediately after forest removal and committed from past deforestation in: (a) all tropical regions combined, (b) the Amazon, (c) the Congo basin, and (d) Southeast Asia, from 1950-1951 through 2009-2010. The error bars represent the 95% confidence interval estimate from up to 100 model iterations.

Even if deforestation had stopped completely by 2010 there were still high amounts of carbon to be released due to past tropical deforestation (Figs. 24 and 25), which we term the 'carbon legacy' of modern tropical deforestation. Naturally, for all regions, in the first few years the amount of carbon released immediately is higher than the one committed for future release, especially in SE Asia due to the higher rates of deforestation estimated in this region in the beginning of the time period covered in our analysis (Fig. 22). However, the committed emissions by the end of the time period (2009), accounted in every region for at least two-thirds of the total carbon emissions.

According to the rates of decay applied in this study and widely used in the community (Houghton 2005; Le Quere et al. 2009; Baccini et al. 2012), it would take seven years for 50% of this carbon legacy to be emitted. From Amazonian deforestation there would be 2.70 Pg C (± 0.70, 95% C.I.) to be released, which is equivalent approximately to seven years of deforestation at the average rate observed in the 2000s (average 0.41 Pg C/yr), whereas from deforestation in the Congo basin there would be 0.82 Pg C (± 0.15, 95% C.I.) to be released, corresponding to ten years of deforestation (average 0.08 Pg C/yr). For SE Asia there would be 0.34 Pg C (± 0.08, 95% C.I.) to be released, corresponding to seven years of deforestation in the region (average 0.05 Pg C/yr). This estimate is the combination of results found for each sub-region (Appendix B - <u>Carbon legacy from Southeast Asia</u>), the fact that it includes areas with very low deforestation such as Papua and Sulawesi, makes the estimate be lower than those found for the Amazon and the Congo basin.

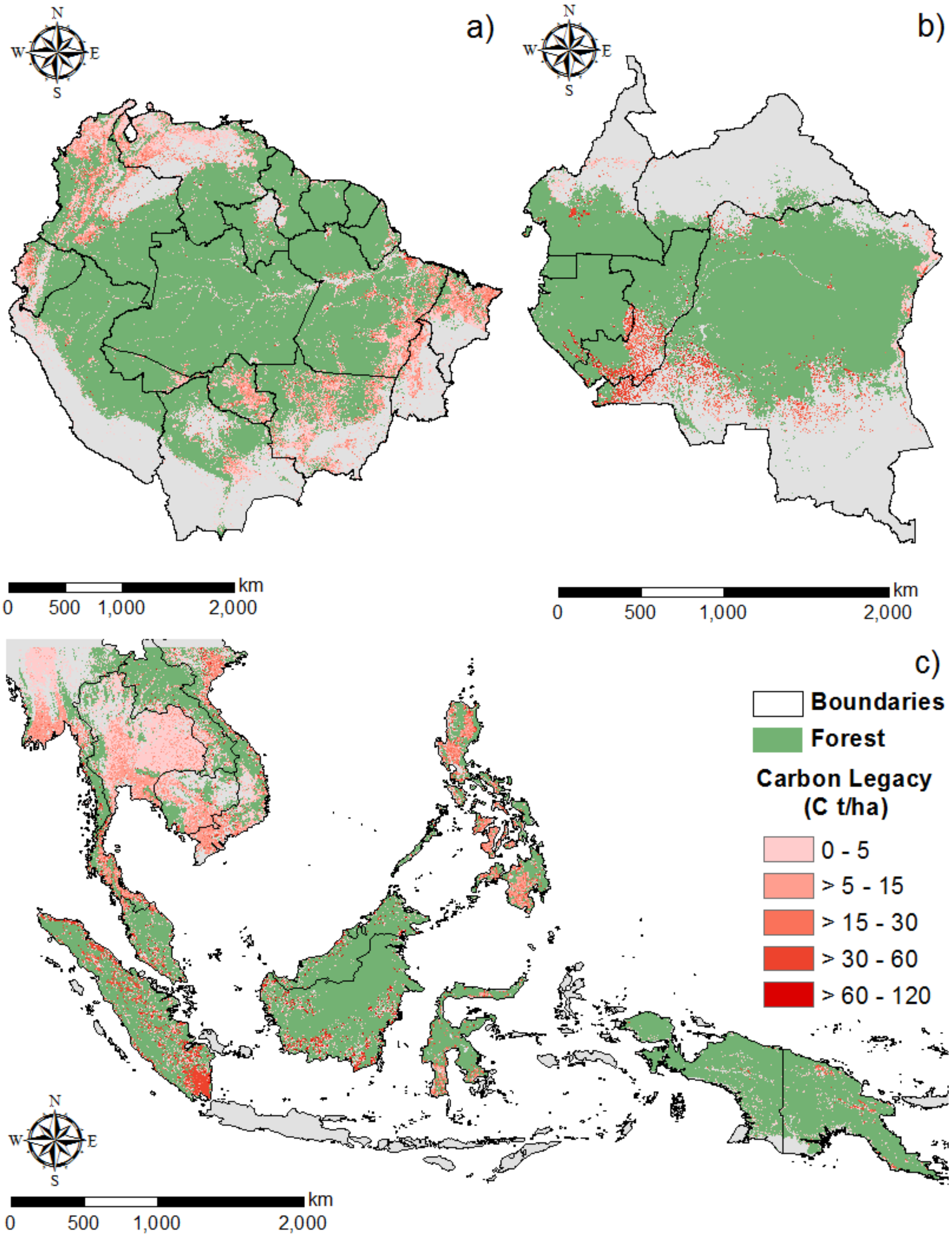


Figure 24 – Carbon left to decay (median) after 2010 (C t/ha) from past tropical deforestation in: (a) the Amazon, (b) the Congo basin and (c) Southeast Asia.

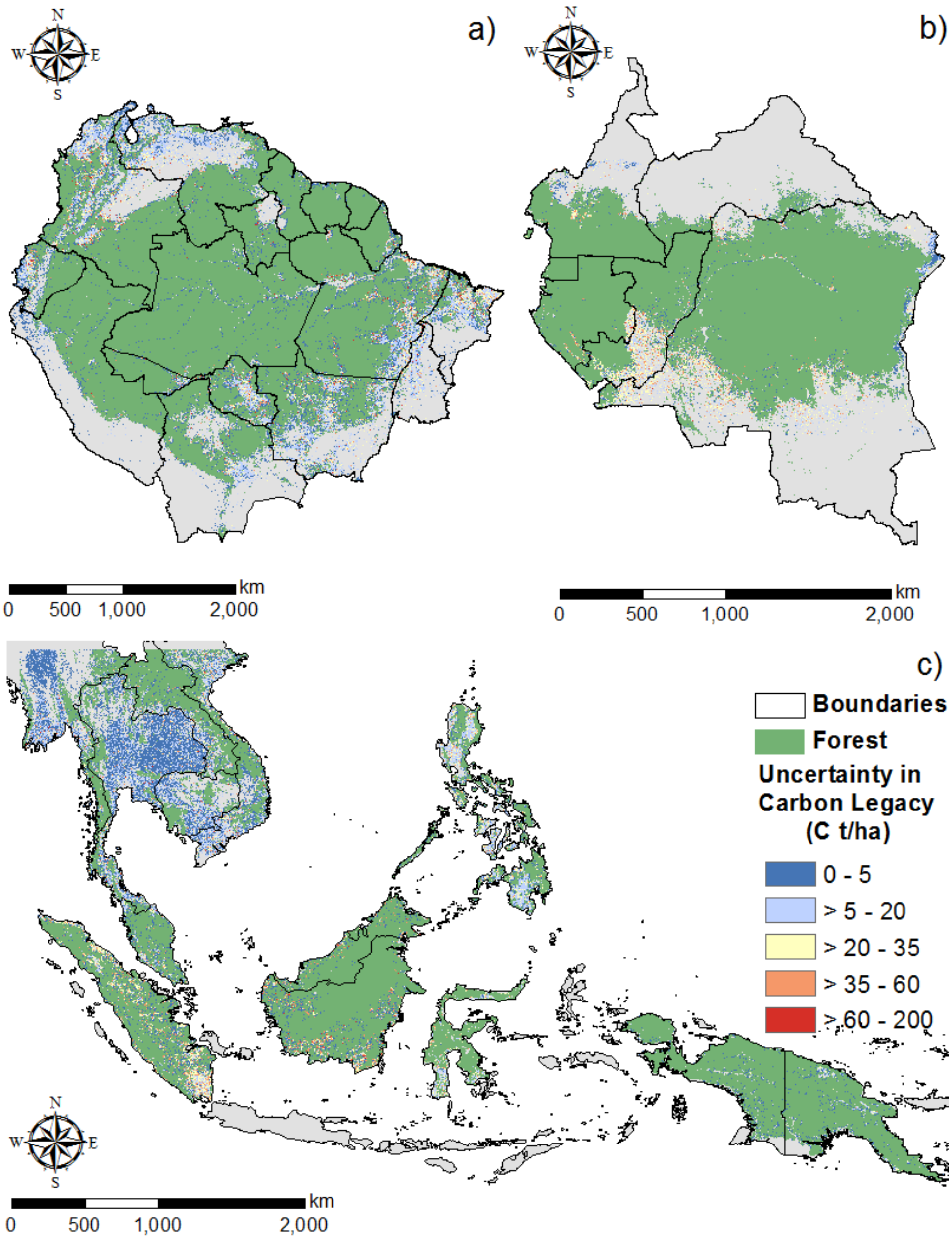


Figure 25 – Uncertainty (Inter-quantile range – IQR) of estimates in the amount of carbon left to decay by 2010 (C t/ha) from past tropical deforestation in: (a) the Amazon, (b) the Congo basin and (c) Southeast Asia.

In the Amazon he carbon legacy would be mainly from the heavily deforested regions of the Arc of deforestation along the south-eastern part of the Amazon (Fig. 24a), whereas in the Congo basin this carbon release would be from the south and east part of the basin where most recent deforestation has happened (Fig. 24b). In SE Asia we can see a clear distinction in the map showing the amount of carbon left to decay, or carbon legacy (Fig. 24c). In the mainland and Philippines archipelago the remaining carbon to decay is a lot less due to the fact that deforestation here has happened further back in time when compared to the other islands, such as Sumatra or Borneo. In these islands the carbon legacy is large due to the high rates of deforestation in the last two decades. The Papua island show little amount of carbon to decay, mainly due to its low deforestation rates, caused among other factors by the island's isolation. The uncertainty found in the carbon legacy estimates are however, much higher in the Congo basin (Fig. 25b) than they are for the Amazon (Fig. 25a) and for most SE Asia (Fig. 25c).

6.4.4) Validation of model spatial predictions

Annual validations on a pixel by pixel basis resulted, as expected, in a weaker predictive power through time of the models in all regions (Fig. 26a, d, g, results for each sub-model of SE Asia are shown in Figs. B2 and B3 in Appendix B). When validated against a time step further from the calibration year, the predictions became more uncertain. By contrast, when we analysed the results cumulatively, model accuracy tended to increase through time although this increase followed a non-linear pattern (Fig. 26a, d, g, plus Figs. B2 and B3 in Appendix B). This is in part explained by the fact that the number of simulations that included 9 time steps to validate gets limited (only models that were calibrated for 2010-2009 can be validated for the full 9 time steps). The best results were achieved in SE Asia which is the region where we used sub-regional models to make the predictions. On an annual basis, omission errors (Fig. 26b, e, h) tended to increase, accompanied by a decrease in the commission errors (Fig. 26c, f, i), especially in SE Asia (Figs. B2 and B3 in Appendix B).

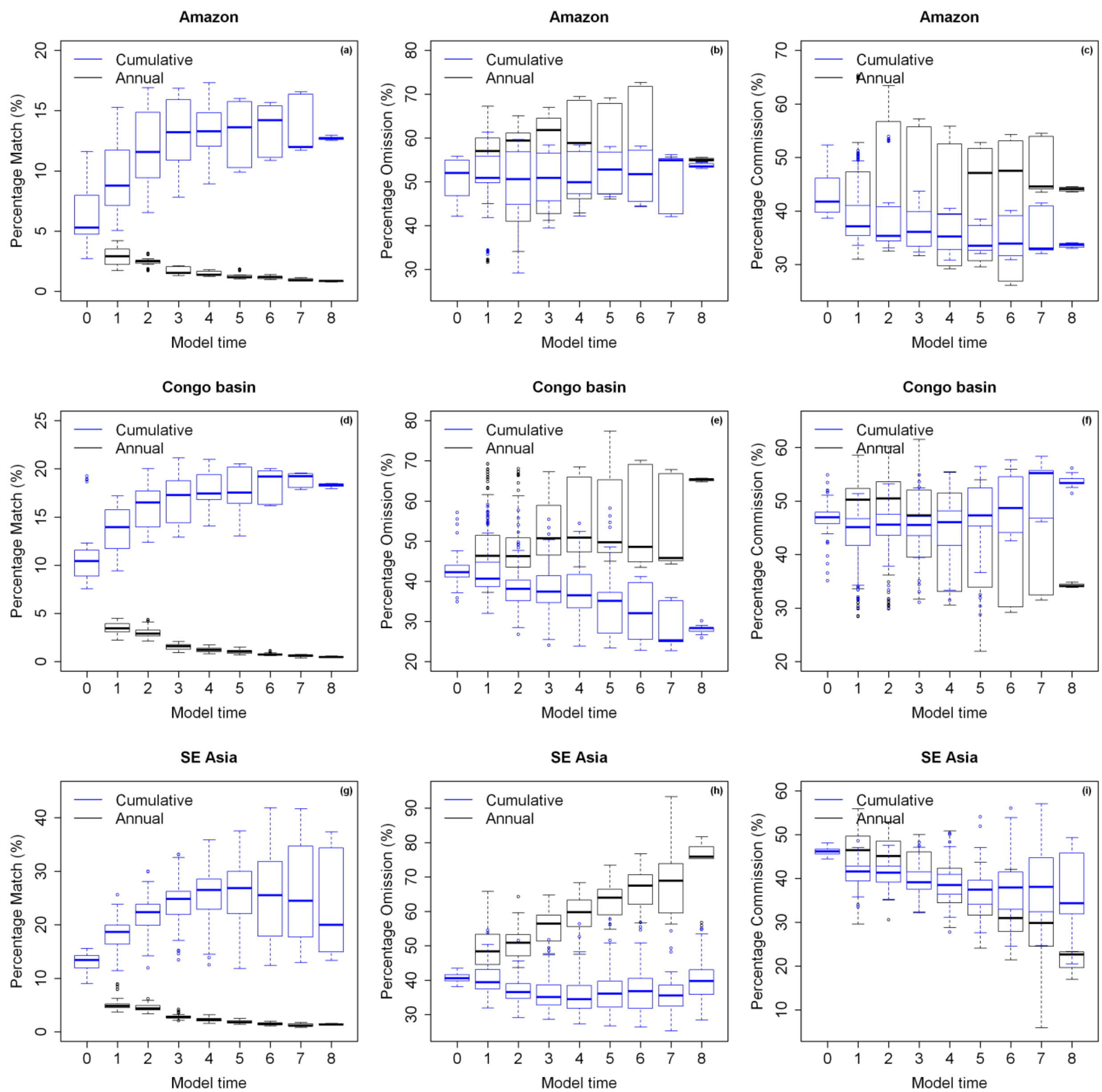


Figure 26 – Pixel by pixel validation results for the first 9 years of model predictions in each three regions: (a), (d), (g) show perfect match; (b), (e), (h) show omission and (c), (f), (i) show commission). The results shown here for SE Asia is the sum of the results found for each sub-model (six in total) applied in this region (results for each of these sub-models can be found in Appendix B, Figs. B2 and B3).

To analyse how well the model captured the spatial pattern of deforestation, and account for near and far misses, we used a distance-based metric to assess how much of the observed events occurred within certain distances of our predictions. As expected, the proportion of observed land cover change increased with the size of the buffer applied to our predictions and this same proportion decreased through time (Fig. 27). This means that predictions made further from the calibration year tended to be more uncertain in concordance with pixel by pixel results. Through time and within 1 km (which corresponds to 1 pixel difference) of our model predictions we were able to capture from 4-9 %, 4-16% and 4-19 % in the Amazon, Congo basin and SE Asia, respectively. This proportion naturally increased with the size of the buffer zone, and we were able to always capture more than 50% of all observed deforestation events through time using a buffer of just 5 km, and more than two thirds when using a buffer of 10 km.

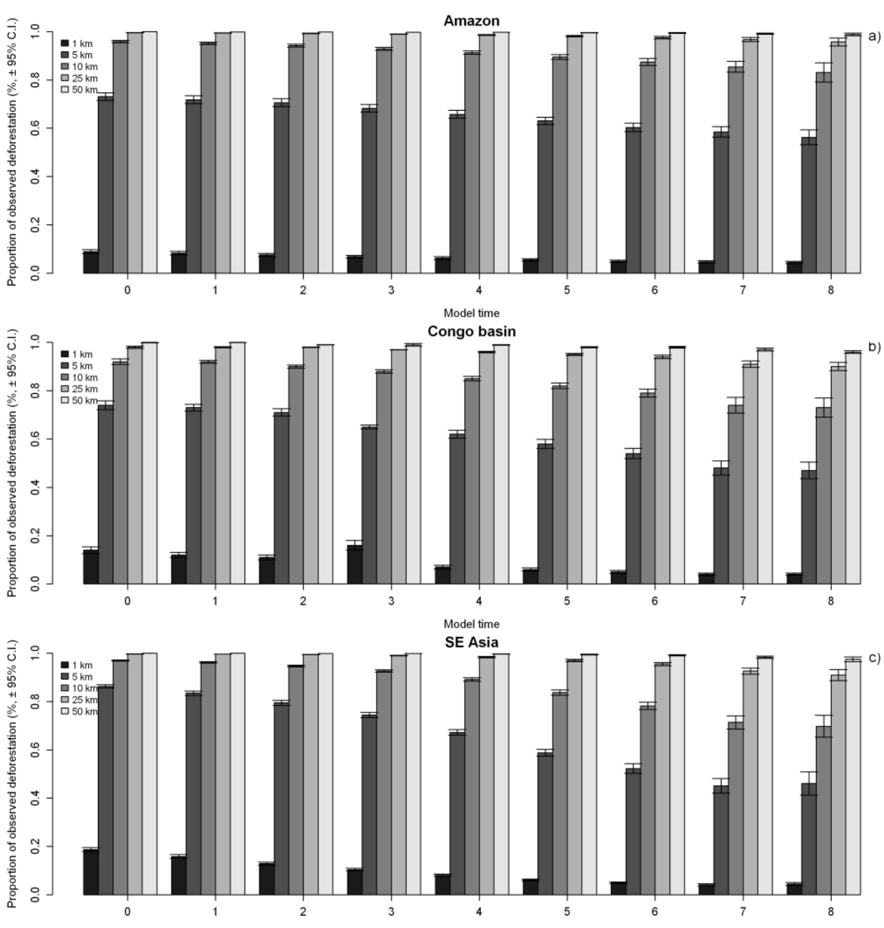


Figure 27 – Distance-based metric of validation: proportion of observed deforestation within 1, 5, 10, 25 and 50 km of model predictions in (a) Amazon, (b) Congo basin and (c) SE Asia through the first 9 years of model predictions (results for each of these sub-models can be found in Appendix B, Figs. B4 and B5).

6.5) Discussion

Although many attempts throughout the years have been made to estimate carbon emissions from tropical deforestation (DeFries et al. 2002; Achard et al. 2004; Baccini et al. 2012), even today, this continues to be a number that is far from certain. This is mainly due to three sources of uncertainties: rate of deforestation, plant carbon content, and processes included in the estimation methodology. Improving confidence in the estimates of carbon emissions from tropical deforestation is vital for policy and conservation.

Using a spatially-explicit model we estimated gross deforestation rates from 1950 through 2009 in the three main tropical regions of the world: the Amazon, the Congo Basin and SE (SE) Asia. Then, combining a bookkeeping carbon model (Houghton 2005) with a terrestrial carbon model (Smith et al. 2012) we estimated resultant carbon emissions. It is important to notice that we only covered the spatial distribution of rainforests rather than all forest extent of the tropics. Further, we assumed that if a forest was cleared in 1950 it will remain deforested until 2009. No regrowth was incorporated in our estimates meaning we do not account for the potential sequestration of carbon that might have occurred after the emissions from deforestation. Similarly, we do not include emissions from logging and soil carbon.

Estimates of deforestation rates found in the literature are highly variable which makes it hard to properly validate them, but we found our model estimates (Fig. 22) to be within the range presented by other authors. Comparing to the results found by Achard et al. (2004) for the 1990s (4.000 – 10.000, 10.000 – 34.000, and 12.000 – 28.000 km^2/yr for the Congo basin, the Amazon and SE Asia, respectively), on average, our estimates were within the range that those found by the authors for all regions. However, if we compare with the results found by DeFries et al. (2002) our estimates are lower for the Congo basin both for the 1980s and 1990s (10.540 – 15.080 and 9.260 – 13.250 km^2/yr, respectively) as well as our estimates for SE Asia in the 1980s (14.910 – 29.290 km^2/yr), but the estimates for the Amazon in both periods (29.160 – 50.850 and 26.240 – 45.740 km^2/yr, respectively) and SE Asia in the 1990s are within range (18.950 – 37.210 km^2/yr). Finally, if we compare our estimates with the reports from FAO (2010) our estimates for the Congo basin are in line for the 1990s but over the reports for the 2000s (~6.000 km^2/yr), whereas for the Amazon and SE Asia our estimates are under the estimates made by FAO in the 1990s and over the estimates for 2000s (~26.000 and 12.000 km^2/yr for SE Asia, ~40.000 and 30.000 km^2/yr for the Amazon, respectively).

Validating our carbon emissions estimates have the same problem which is that different research groups use various methodologies and incorporate several processes in their estimates. Incorporating all past emissions since 1950 we were expecting to find our estimates to be much higher than those presented by other authors. However, including uncertainty in the plant carbon content resulted in a lower estimate of biomass when compared to two other datasets (Baccini *et al.*, 2012, Saatchi *et al.*, 2011). This combination prevented our results from being higher than those presented in the literature. Our estimates for the Congo basin were found to be lower than the estimates made by Baccini et al. (2012) (0.23 Pg C/yr) and Harris et al. (2012a) (0.11 Pg C/yr); a pattern also observed for the Amazon (0.41 versus 0.47 and 0.44 Pg C/yr, respectively); but not for SE Asia where our estimates were found to be higher (0.30 versus 0.11 and 0.26 Pg C/yr, respectively).

When compared to the estimates for the 1990s found by Achard et al. (2004) our results are significantly lower (1.07 ± 0.3 *versus* 0.59 ± 0.03 Pg C/yr, respectively), however we do not include emissions from soil carbon or degradation which would inevitably lead to higher amounts of carbon being released. Further, the author does not use a spatial dataset of biomass, instead an average biomass value is used which can lead to overestimation. Our estimates are however within the range (0.5–1.4 Pg C/yr) predicted by DeFries et al. (2002) for the same time period. The same is true when we compare our results (0.44 ± 0.03 Pg C/yr) with those predicted by (DeFries et al. 2002) for the 1980s (0.3–0.8 Pg C/yr). Although we calculate cumulative losses since the 1950s, we emphasise that we do not include soil carbon losses as well as degradation which will necessarily lead to an underestimation of our results. Unfortunately, there are no systematic datasets on soil carbon losses from deforestation or logging at a global scale to be able to include rigorously in our estimates. If we add 25-35% to our estimates due to logging (Houghton et al. 2012) than our estimates would be higher (1.2 - 1.3 Pg C/yr) that all other estimates for 2009. This is due to the fact that previous estimates do not include the carbon legacy of emissions from past deforestation.

Emissions from the burning of fossil fuels are the major source of carbon to the atmosphere at a global scale (Houghton 2005; Le Quere et al. 2009; Baccini et al. 2012). However, when we analyse the tropics only this relationship inverts with emissions from land cover change being responsible for the majority of carbon released to the atmosphere. In particular, using the 2012 dataset on fuel combustion emission by country (1971-2009) from the International Energy Agency (IEA, 2013) we compared our estimates of carbon dioxide released from land cover change with their fossil fuel emissions reports. We found that land cover change emissions were three to two times higher than those by the combustion of fuel within the tropics. The difference between the two types of emissions is getting smaller in recent years mainly due to the rapid economic growth of Brazil and some countries of SE Asia. The discrepancy between emissions is particularly accentuated for the Congo basin where the ratio is up to 60 times higher due to the very low quantities of emissions from fossil fuels. In SE Asia, on the contrary, emissions from land cover change are now almost at the same level as emissions from fossil fuel combustion whereas in the 1970s the former were four times higher than the later.

Under the Kyoto protocol all countries covered in our study were included as "Developing countries without binding targets", which means that they had no obligation to meet any target related to emissions (UNFCCC, 2013). However, in 2010, under the Cancun Climate Change Conference from the United Nations Framework Convention on Climate Change (UNFCCC) some of these countries including Brazil and Indonesia voluntarily committed to reduce their emissions. While Brazil aimed to reduce its emissions by 21-25% from 2005 levels, Indonesia aimed to reduce its emissions by at least 26% by 2020 (WRI, 2013). Our results show that even if deforestation in the tropics had stopped completely in 2010, there were still large quantities of carbon to be released to the atmosphere and this legacy of past deforestation will ensure that reductions in carbon emissions will lag behind reductions in deforestation rates. Given that the great majority of carbon emissions from these countries come from the change in land cover, the effort required to reduce emissions may be higher than expected once this carbon legacy is accounted for.

6.6) Acknowledgments

This work was funded by the Grantham Institute for Climate Change, Microsoft Research Cambridge and the European Research Council.

Chapter 7 – Discussion and Conclusions

7.1) Summary of work

The aim here is to provide a short summary of the work performed in this thesis while exploring the linkages between the data chapters presented above. The quantitative review of land use and land cover change models in the Amazon (Chapter 3) showed a few shortcomings of current models applied in this region. These included the lack of openness with regard to presenting and making available the model inputs, code and outputs; a disregard for the importance of validating model outputs; and in most cases the absence of uncertainty associated with predictions. These features restrain models to have a stronger influence over public perceptions of tropical land cover change and its consequences. Further, it prevents their use in policy decisions at all political levels. The model developed in this thesis (Chapter 4) aimed to address some of these points by making available all the data used in the process of calibrating and validating the model; by performing stringent spatially-explicit validations at multiple time steps; and by explicitly including uncertainty around all predictions – amount and location of change. These aspects facilitate a critical view of the model by decision-makers and allow the comparison of model prediction for the same region between different models. Additionally, it allow researchers to build ensembles of competing model predictions which is as of yet a gap in the land cover change community as highlighted in Chapter 3.

When analysing the temporal dynamics of deforested patch sizes in the Brazilian Amazon (Chapter 2), two distinct time periods of rising and decreasing deforestation rates and of absence and presence of actions by the Brazilian government in preventing deforestation were identified. These periods of time were taken into consideration while choosing the years to calibrate the model developed in Chapter 4. In particular, two scenarios of deforestation for the same region were simulated and the model was tested using two datasets of years that fell inside each of the time periods referred above. This resulted in very different estimates of deforestation by 2050 both temporally and spatially. As expected, the scenario that used data for the period before 2004 (no environmental policies implemented) revealed very high deforestation rates that would consume the most accessible forest quickly, and then deforestation rates would slow down mainly due to a lack of accessibility. In contrast, the scenario that used data for the period after 2004 (environmental policies started to be implemented) revealed a much lower but steadier deforestation rate.

These results encouraged the interest in exploring the role that calibration year played in the model's performance. As such, in Chapter 5 apart from updating the model in order to incorporate two more land cover transitions (deforested→regeneration and regeneration→deforested, adding to the existing forest→deforested), the temporal dynamics of LCC drivers was explored as well as the role of the year chosen for model calibration, validation, transition length and time step. Using a large historical database of land cover in Machadinho d'Oeste, Rondonia, it was found that model calibration year is of extreme importance for model predictability and performance. Therefore, in the final data chapter of this thesis (Chapter 6) a new feature was included in the model to remove the dependency from a single calibration time period to make predictions of land cover change. Acknowledging that global land cover change datasets have large uncertainties attached, instead of relying on a single year for model calibration, a set of years was used stochastically. This means that a set of model predictions was built using several starting points, which allowed us to provide uncertainty in both deforestation and carbon estimates (Chapter 6). Despite the model being already ready to include several land cover transitions and the dataset used in this chapter provided information for many land cover types this was avoided. There are still very high uncertainties attached to global datasets of land cover, which would increase greatly the risk of modelling the wrong class in the wrong location. These errors were minimized by reducing the number of land cover transitions in the model to the minimum (forest→deforested).

Finally, results from the model simulations in the Brazilian Amazon (Chapter 4) revealed that the model works well in predicting both the rate and location of change, although, naturally this becomes more uncertain, on an annual basis, through time. Additionally, it was demonstrated that it is possible to use the model to simulate different scenarios based on distinct initial settings as exemplified by the two scenarios (before and after the implementation of conservation policies by the Brazilian government) of deforestation between 2002 and 2050 were run for the Brazilian Amazon (Chapter 4). These predictions from different scenarios can be combined with other datasets to determine, among other things, carbon and biodiversity losses. In this thesis, the applicability of the model predictions was illustrated at the global scale (three main tropical regions: Amazon, Central Africa and Southeast Asia) by determining the amount of carbon loss from 1950 through 2010, which was names the ´carbon legacy´. This was achieved by combining multiple sources of information – the model developed in this thesis, a terrestrial carbon model and a widely used bookkeeping carbon model. Results showed that even if deforestation had stopped completely by 2010 there was still a very large amount of carbon to be released over the next years through the decay of remaining biomass. This analysis shows that human activity have already had a great impact on tropical forests and in the carbon cycle which urges the implementation of conservation policies and actions in order to prevent further depletion of these ecosystems.

7.2) Outputs

All input data used in this thesis as well as the C^{++} code to run the spatial model, and all model outputs described in each chapter are available upon request via e-mail (i.rosa09@imperial.ac.uk or isamdr@gmail.com). Model outputs include the annual probability maps of land cover change (Chapter 4, 5 and 6), videos showing the cumulative deforestation probability from 2002 through 2050 in the Brazilian Amazon from Chapter 4 and the cumulative loss of carbon in each tropical region from Chapter 5.

7.3) Directions for future research

In this thesis a model of land cover change was developed in order to generate probability maps of future tropical deforestation and estimate the annual rate of change in forest cover (Chapter 4). The predictions of deforestation were stochastic, allowing for uncertainty in the parameters used to predict deforestation and uncertainty in the exact location of deforestation events. The modelling approach was also adapted to back-cast the rate and pattern of deforestation, and estimate carbon emissions in the three main tropical regions of the globe (Amazon, Congo basin and Africa) (Chapter 6). Further it was demonstrated that this approach can predict multiple LUC transitions (*e.g.* forest to agriculture, agriculture to abandoned land) (Chapter 5).

In the perspective of the model here developed, there is room for improvement as these were just the first steps on its development. Some chains of research that could be followed, in order to keep improving the model and make it flexible enough to manage the dynamics of real world to provide better estimates and better predictions, include: multi-scale parameterisation and predictions; incorporation other land cover types, thus incorporating other processes of land cover change (*e.g.* fire); dynamic links with other models such road network model and climate change models; among others.

The methods already developed provide a solid base from which global predictions of land cover change could be modelled using a multi-scale dynamic approach, but working at this scale would require substantial additions to the modelling framework. For instance, region and land use transition-specific parameterised sub-models would be needed to cover the diversity of terrestrial environments and ecosystems. To address the different drivers of land cover change in the different biomes (*e.g.* tropical deforestation *versus* urbanization *versus* natural fires in savannahs) a large set of sub-models would need to be parameterised and updated dynamically through time. Furthermore, the probability of each specific land cover transition occurring in a particular part of the world would be modelled as a function of a suite of potential drivers, including variables such as proximity to roads, socio-economic status, topography and climate. These parameter estimates would then be interpolated across space to identify spatial discontinuities in the different processes underlying the spatial patterns of land cover change, and thereby delineating the spatial boundaries that would define the extent of each sub-model. The sub-models would run semi-independently of each other, using their own unique set of parameter estimates, but retaining spatial linkages to ensure that changes in one environment can impact the rate and pattern of changes in adjacent political and/or environmental units. An additional benefit of generating parameter estimates at a range of spatial scales is that we can create a final model that can be used to predict land cover change at any spatial scale on any part of the globe, ensuring that the outputs could be integrated into carbon, climate and biodiversity models regardless of their spatial resolution.

These improvements would make the model more informative and helpful to potentially be used in conservation decisions at the policy level which, unfortunately, does not happen yet. Although land cover change processes are relatively easy for the general public to comprehend (*e.g.* tropical forest being replaced by pastures), and share the same importance as climate in the future of our changing world, land cover change models have not yet the same impact as other models (*e.g.* climate change) have to influence public perception and/or public policies. The lack of consistent and standardised outputs that could be used to make model comparisons essential for a critical view of the problem might help explaining why that happens. For that, as raised in one of the chapters of this thesis (Chapter 3), changes are needed in the discipline, especially in three aspects: improving the transparency of modelling frameworks, sharing model outputs, and consistently validate these same outputs.

7.4) Conclusions

Human driven environmental changes such as global land use change (LUC), driven by rapid human population growth and increasing demand for agricultural and forest products (Foley et al. 2005), are impacting the balance of the Earth system. LUC is now one of the major causes of biodiversity loss (Van Vuuren et al. 2006), the second largest source of carbon emissions to the atmosphere (Le Quere et al. 2009), and has a major impact on the water cycle (Sterling et al. 2012). In this thesis a probabilistic model that determined the magnitude and location of this change was developed, explicitly associating a measure of uncertainty around its predictions. This uncertainty is fundamental to weight different model predictions which enable policy-makers to critically analyse different model outputs. Knowing where and at what rate LUC will continue in the future would allow us to anticipate its impacts and develop preventive and/or adaptive measures. Furthermore, because predictions of land cover change influence the outcomes of many different branches of science (*e.g.* biodiversity, carbon) it is ever more important that model outputs are available and the accuracy and uncertainty of these predictions is clear.

References

Abdullah, S., and A. Hezri. 2008. From Forest Landscape to Agricultural Landscape in the Developing Tropical Country of Malaysia: Pattern, Process, and Their Significance on Policy. Environmental Management **42**:907-917.

Abdullah, S. A., and N. Nakagoshi. 2007. Forest fragmentation and its correlation to human land use change in the state of Selangor, peninsular Malaysia. Forest Ecology and Management **241**:39-48.

Achard, F., H. D. Eva, P. Mayaux, H.-J. Stibig, and A. Belward. 2004. Improved estimates of net carbon emissions from land cover change in the tropics for the 1990s. Global Biogeochem. Cycles **18**:GB2008.

Aguiar, A. P. D., G. Camara, and M. I. S. Escada. 2007. Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. Ecological Modelling **209**:169-188.

Ahmed, S., C. Souza, Jr., J. Riberio, and R. Ewers. 2013. Temporal patterns of road network development in the Brazilian Amazon. Regional Environmental Change:1-11.

Ahmed, S., C. Souza Jr, J. Riberio, and R. Ewers. 2012. Temporal patterns of road network development in the Brazilian Amazon. Regional Environmental Change:in press.

Alcorn, J. B. 1993. Indigenous peoples and conservation. Conservation Biology **7**:424 - 426.

Alston, L. J., G. D. Libecap, and S. R. 1996. The Determinants and Impact of Property Rights: Land Titles on the Brazilian Frontier. National Bureau of Economic Research, Inc.

Alves, D. S. 2002. Space-time dynamics of deforestation in Brazilian Amazonia. International Journal of Remote Sensing **23**:2903-2908.

Angelsen, A., and D. Kaimowitz. 1999. Rethinking the causes of deforestation: Lessons from economic models. World Bank Research Observer **14**:73-98.

Araújo, A. C., A. D. Nobre, B. Kruijt, J. A. Elbers, R. Dallarosa, P. Stefani, C. von Randow, A. O. Manzi, A. D. Culf, J. H. C. Gash, R. Valentini, and P. Kabat. 2002. Comparative measurements of carbon dioxide fluxes from two nearby towers in a central Amazonian rainforest: The Manaus LBA site. J. Geophys. Res. **107**:8090.

Araujo, C., C. A. Bonjean, J. L. Combes, P. C. Motel, and E. J. Reis. 2009. Property rights and deforestation in the Brazilian Amazon. Ecological Economics **68**:2461-2468.

Arima, E., R. Walker, S. Perz, and M. Caldas. 2005a. Loggers and Forest Fragmentation: Behavioral Models of Road Building in the Amazon Basin. Annals of the Association of American Geographers **95**:525-541.

Arima, E. Y., R. T. Walker, S. G. Perz, and M. Caldas. 2005b. Loggers and Forest Fragmentation: Behavioral Models of Road Building in the Amazon Basin. Annals of the Association of American Geographers **95**:525-541.

Arima, E. Y., R. T. Walker, M. Sales, C. Souza Jr., and S. G. Perz. 2008a. The Fragmentation of Space in the Amazon Basin: Emergent Road Networks. Photogrammetric Engineering & Remote Sensing **74**:699–709.

Arima, E. Y., R. T. Walker, M. Sales, C. Souza Jr., and S. G. Perz. 2008b. The Fragmentation of Space in the Amazon Basin: Emergent Road Networks. Photogrammetric Engineering & Remote Sensing **74**:699 - 709.

Armenteras, D., G. Rudas, N. Rodriguez, S. Sua, and M. Romero. 2006. Patterns and causes of deforestation in the Colombian Amazon. Ecological Indicators **6**:353-368.

Asner, G., J. Kellner, T. Kennedy-Bowdoin, D. Knapp, C. Anderson, and e. al. 2013. Forest Canopy Gap Distributions in the Southern Peruvian Amazon. PLOS One **8**:e60875.

Asner, G. P., E. N. Broadbent, P. J. C. Oliveira, M. Keller, D. E. Knapp, and J. N. M. Silva. 2006. Condition and fate of logged forests in the Brazilian Amazon. Proceedings of the National Academy of Sciences of the United States of America **103**:12947-12950.

Asner, G. P., T. K. Rudel, T. M. Aide, R. Defries, and R. Emerson. 2009. A Contemporary Assessment of Change in Humid Tropical Forests. Conservation Biology **23**:1386-1395.

Assunção, J., C. Gandour, and R. Rocha. 2012. Deforestation Slowdown in the Legal Amazon: Prices or Policies? Climate Policy Initiative.

Baccini, A., S. J. Goetz, W. S. Walker, N. T. Laporte, M. Sun, D. Sulla-Menashe, J. Hackler, P. S. A. Beck, R. Dubayah, M. A. Friedl, S. Samanta, and R. A. Houghton. 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. Nature Climate Change:182–185.

Baccini, A., N. Laporte, S. J. Goetz, M. Sun, and H. Dong. 2008. A first map of tropical Africa's above-ground biomass derived from satellite imagery. Environmental Research Letters **3**:045011.

Barlow, J., R. M. Ewers, L. Anderson, L. E. O. C. Aragao, T. R. Baker, E. Boyd, T. R. Feldpausch, E. Gloor, A. Hall, Y. Malhi, W. Milliken, M. Mulligan, L. Parry, T. Pennington, C. A. Peres, O. L. Phillips, R. M. Roman-Cuesta, J. A. Tobias, and T. A. Gardner. 2011. Using learning networks to understand complex systems: a case study of biological, geophysical and social research in the Amazon. Biological Reviews **86**:457-474.

Barona, E., N. Ramankutty, G. Hyman, and O. T. Coomes. 2010. The role of pasture and soybean in deforestation of the Brazilian Amazon. Environmental Research Letters **5**:024002.

Barreto, P., A. Pinto, B. Brito, and S. Hayashi. 2008. Quem é dono da Amazônia? Uma análisa do recadastramento de imóveis rurais. . Instituto do Homem e Meio Ambiente da Amazônia, Belém, Brazil.

Betts, R. A., Y. Malhi, and J. T. Roberts. 2008. The future of the Amazon: new perspectives from climate, ecosystem and social sciences. Philosophical Transactions of the Royal Society B: Biological Sciences **363**:1729-1735.

Boakes, E. H., G. M. Mace, P. J. K. McGowan, and R. A. Fuller. 2010. Extreme contagion in global habitat clearance. Proceedings of the Royal Society B: Biological Sciences **277**:1081-1085.

Bonan, G. B. 2008. Forests and climate change: Forcings, feedbacks, and the climate benefits of forests. Science **320**:1444-1449.

Borchert, R. 1998. Responses of tropical trees to rainfall seasonality and its long-term changes. Climatic Change **39**:381-393.

Brandão, A. O., and C. M. Souza. 2006. Mapping unofficial roads with Landsat images: a new tool to improve the monitoring of the Brazilian Amazon rainforest. International Journal of Remote Sensing **27**:177 - 189.

Camara, G., D. M. Valeriano, and J. V. Soares. 2006. Metodologia para o Cálculo da Taxa Anual de Desmatamento na Amazônia Legal. Instituto Nacional de Pesquisas Espaciais, São José dos Campos, São Paulo, Brazil

Cantelaube, P., and J. M. Terres. 2005. Seasonal weather forecasts for crop yield modelling in Europe. Tellus **57**:476-487.

Carlson, K. M., L. M. Curran, D. Ratnasari, A. M. Pittman, B. S. Soares-Filho, G. P. Asner, S. N. Trigg, D. A. Gaveau, D. Lawrence, and H. O. Rodrigues. 2012. Committed carbon emissions, deforestation, and community land conversion from oil palm plantation expansion in West Kalimantan, Indonesia. Proceedings of the National Academy of Sciences **109**:7559-7564.

Carpentier, C. L., S. A. Vosti, and J. Witcover. 2000. Intensified production systems on western Brazilian Amazon settlement farms: could they save the forest? Agriculture Ecosystems & Environment **82**:73-88.

Carreiras, J. M. B., J. B. Melo, R. M. Lucas, and Y. E. Shimabukuro. 2012. Assessing regrowth and deforestation processes in western Amazonia (Rondônia, Brazil) using high frequency time series of high spatial resolution remote sensing data. SPIE Remote Sensing 2012, Edinburgh, UK.

Chambers, J. Q., N. Higuchi, J. P. Schimel, L. V. Ferreira, and J. M. Melack. 2000. Decomposition and carbon cycling of dead trees in tropical forests of the central Amazon. Oecologia **122**:380 – 388.

Chomitz, K., and C. Griffiths. 1996. Deforestation, shifting cultivation, and tree crops in Indonesia: nationwide patterns of smallholder agriculture at the forest frontier. Environment, Infrastructure, and Agriculture Division, Policy Research Department, The World Bank, Washington, USA.

Chomitz, K. M., and T. S. Thomas. 2001. Geographic Patterns of Land Use and Land Intensity in the Brazilian Amazon. World Bank Policy Research Working Paper **2687**.

Chomitz, K. M., and T. S. Thomas. 2003. Determinants of land use in Amazonia: A fine-scale spatial analysis. American Journal of Agricultural Economics **85**:1016-1028.

Chuvieco, E. 1993. Integration of linear programming and GIS for land-use modelling. International Journal of Geographical Information Systems **7**:71-83.

Clark Labs. 2007. The Land Change Modeler for Ecological Sustainability. IDRISI focus paper, http://www.clarklabs.org/applications/upload/Land-Change-Modeler-IDRISI-Focus-Paper.pdf.

Colchester, M., N. J. Andiko, M. Sirait, A. Y. Firdaus, A. Surambo, and H. Pane. 2006. Promised land: palm oil and land acquisition in Indonesia—implications for local communities and indigenous peoples, Moreton-in-Marsh, UK.

Costa, M. H., A. Botta, and J. A. Cardille. 2003. Effects of large-scale changes in land cover on the discharge of the Tocantins River, Southeastern Amazonia. Journal of Hydrology **283**:206-217.

Costa, M. H., and J. A. Foley. 2000. Combined Effects of Deforestation and Doubled Atmospheric CO2 Concentrations on the Climate of Amazonia. Journal of Climate **13**:18-34.

Cox, P. M., R. A. Betts, M. Collins, P. P. Harris, C. Huntingford, and C. D. Jones. 2004. Amazonian forest dieback under climate-carbon cycle projections for the 21st century. Theoretical and Applied Climatology **78**:137-156.

Cox, P. M., R. A. Betts, C. D. Jones, S. A. Spall, and I. J. Totterdell. 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. Nature **408**:184-187.

Dale, V. H., R. V. Oneill, F. Southworth, and M. Pedlowski. 1994. Modeling Effects of Land Management in the Brazilian Amazonian Settlement of Rondonia. Conservation Biology **8**:196-206.

Davies, R. G., C. D. L. Orme, V. Olson, G. H. Thomas, S. G. Ross, T.-S. Ding, P. C. Rasmussen, A. J. Stattersfield, P. M. Bennett, T. M. Blackburn, I. P. F. Owens, and K. J. Gaston. 2006. Human impacts and the global distribution of extinction risk. Proceedings of the Royal Society B: Biological Sciences **273**:2127-2133.

de Koning, G. H. J., A. Veldkamp, and L. O. Fresco. 1999a. Exploring changes in Ecuadorian land use for food production and their effects on natural resources. Journal of Environmental Management **57**:221-237.

de Koning, G. H. J., P. H. Verburg, A. Veldkamp, and L. O. Fresco. 1999b. Multi-scale modelling of land use change dynamics in Ecuador. Agricultural Systems **61**:77-93.

de Wasseige, P., N. de Marcken, F. Bayol, P. Hiol Hiol, B. Mayaux, R. Desclée, A. Nasi, P. Billand, R. Defourny, and E. a. Atyi. 2012. The Forests of the Congo Basin – State of the Forest 2010 European Union Publications Office

Deadman, P., D. Robinson, E. Moran, and E. Brondizio. 2004. Colonist household decisionmaking and land-use change in the Amazon Rainforest: an agent-based simulation. Environment and Planning B-Planning & Design **31**:693-709.

DeFries, R. S., R. A. Houghton, M. C. Hansen, C. B. Field, D. Skole, and J. Townshend. 2002. Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s. Proceedings of the National Academy of Sciences **99**:14256-14261.

Duveiller, G., P. Defourny, B. Desclée, and P. Mayaux. 2008. Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts. Remote Sensing of Environment **112**:1969-1981.

Eastman, J. R., M. E. Van Fossen, and L. A. Solarzano. 2005. Transition potential modeling for land cover change. . Pages 357 - 386 in D. Maguire, M. Batty, and M. Goodchild, editors. GIS, Spatial Analysis and Modeling. ESRI Press, Redlands, CA.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann, J. Li, L. G. Lohmann, B. A. Loiselle, G. Manion, C. Moritz, M. Nakamura, Y. Nakazawa, J. McC. M. Overton, A. Townsend Peterson, S. J. Phillips, K. Richardson, R. Scachetti-Pereira, R. E. Schapire, J. Soberón, S. Williams, M. S. Wisz, and N. E. Zimmermann. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography **29**:129-151.

Escada, M. I. S., A. M. V. Monteiro, A. P. D. Aguiar, T. G. S. Carneiro, and G. Camara. 2005. Análise de padrões e processos de ocupação para a construção de modelos na Amazônia: experimentos em Rondônia. Pages 2973-2983 in INPE, editor. Simpósio Brasileiro de Sensoriamento Remoto. INPE, Goiânia, Brazil.

Etter, A., C. McAlpine, S. Phinn, D. Pullar, and H. Possingham. 2006a. Unplanned land clearing of Colombian rainforests: Spreading like disease? Landscape and Urban Planning **77**:240-254.

Etter, A., C. McAlpine, K. Wilson, S. Phinn, and H. Possingham. 2006b. Regional patterns of agricultural land use and deforestation in Colombia. Agriculture Ecosystems & Environment **114**:369-386.

Evans, T. P., A. Manire, F. de Castro, E. Brondizio, and S. McCracken. 2001. A dynamic model of household decision-making and parcel level landcover change in the eastern Amazon. Ecological Modelling **143**:95-113.

Ewers, R. M., and W. F. Laurance. 2006. Scale-dependent patterns of deforestation in the Brazilian Amazon. Environmental Conservation **33**:203-211.

Ewers, R. M., W. F. Laurance, and C. M. Souza. 2008. Temporal fluctuations in Amazonian deforestation rates. Environmental Conservation **35**:303-310.

Fa, J. E., C. A. Peres, and J. Meeuwig. 2002. Bushmeat exploitation in tropical forests: an intercontinental comparison. Conservation Biology **16**:232-237.

FAO. 2010. Global forest resources assessment 2010 - main report. Food and Agriculture Organization of the United Nations, Rome.

Fearnside, P. M. 1993. Deforestation in Brazilian Amazonia - the Effect of Population and Land-Tenure. Ambio **22**:537-545.

Fearnside, P. M. 2005. Deforestation in Brazilian Amazonia: History, rates, and consequences. Conservation Biology **19**:680-688.

Fearnside, P. M. 2008. The Roles and Movements of Actors in the Deforestation of Brazilian Amazonia. Ecology and Society **13**:-.

Feddema, J. J., K. W. Oleson, G. B. Bonan, L. O. Mearns, L. E. Buja, G. A. Meehl, and W. M. Washington. 2005. The Importance of Land-Cover Change in Simulating Future Climates. Science **310**:1674-1678.

Ferraz, S. F. D., C. A. Vettorazzi, D. M. Theobald, and M. V. R. Ballester. 2005. Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondonia, Brazil: assessment and future scenarios. Forest Ecology and Management **204**:67-83.

Finer, M., C. N. Jenkins, S. L. Pimm, B. Keane, and C. Ross. 2008. Oil and Gas Projects in the Western Amazon: Threats to Wilderness, Biodiversity, and Indigenous Peoples. PLoS ONE **3**:e2932.

Foley, J. A., G. P. Asner, M. H. Costa, M. T. Coe, R. DeFries, H. K. Gibbs, E. A. Howard, S. Olson, J. Patz, N. Ramankutty, and P. Snyder. 2007. Amazonia revealed: forest degradation and loss of ecosystem goods and services in the Amazon Basin. Frontiers in Ecology and the Environment **5**:25-32.

Foley, J. A., M. H. Costa, C. Delire, N. Ramankutty, and P. Snyder. 2003. Green surprise? How terrestrial ecosystems could affect earth's climate. Frontiers in Ecology and the Environment **1**:38-44.

Foley, J. A., R. DeFries, G. P. Asner, C. Barford, G. Bonan, S. R. Carpenter, F. S. Chapin, M. T. Coe, G. C. Daily, H. K. Gibbs, J. H. Helkowski, T. Holloway, E. A. Howard, C. J. Kucharik, C. Monfreda, J. A. Patz, I. C. Prentice, N. Ramankutty, and P. K. Snyder. 2005. Global Consequences of Land Use. Science **309**:570-574.

Forman, R. T. T., and L. E. Alexander. 1998. Roads and Their Major Ecological Effects. Annu. Rev. Ecolo. Syst. **29**:207 - 231.

Friedl, M. A., D. K. McIver, J. C. F. Hodges, X. Y. Zhang, D. Muchoney, A. H. Strahler, C. E. Woodcock, S. Gopal, A. Schneider, A. Cooper, A. Baccini, F. Gao, and C. Schaaf. 2002. Global land cover mapping from MODIS: algorithms and early results. Remote Sensing of Environment **83**:287-302.

Frohn, R. C., K. C. McGwire, V. H. Dale, and J. E. Estes. 1996. Using satellite remote sensing analysis to evaluate a socio-economic and ecological model of deforestation in Rondv¥nia, Brazil. International Journal of Remote Sensing **17**:3233-3255.

Fujisaka, S., W. Bell, N. Thomas, L. Hurtado, and E. Crawford. 1996. Slash-and-burn agriculture, conversion to pasture, and deforestation in two Brazilian Amazon colonies. Agriculture, Ecosystems & Environment **59**:115-130.

Funk, S., M. Salathé, and V. A. A. Jansen. 2010. Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of The Royal Society Interface **7**:1247-1256.

Galford, G. L., J. M. Melillo, D. W. Kicklighter, T. W. Cronin, C. E. P. Cerri, J. F. Mustard, and C. C. Cerri. 2010. Greenhouse gas emissions from alternative futures of deforestation and agricultural management in the southern Amazon. Proceedings of the National Academy of Sciences.

Garcia, R. A., B. S. Soares, and D. O. Sawyer. 2007. Socioeconomic dimensions, migration, and deforestation: An integrated model of territorial organization for the Brazilian Amazon. Ecological Indicators **7**:719-730.

Gates, W. L. 1992. AMIP: the atmospheric model intercomparison project. Bulletin of the American Meteorological Society **73**:1962-1970.

Gaveau, D. L. A., J. Epting, O. Lyne, M. Linkie, I. Kumara, M. Kanninen, and N. Leader-Williams. 2009a. Evaluating whether protected areas reduce tropical deforestation in Sumatra. Journal of Biogeography **36**:2165-2175.

Gaveau, D. L. A., M. Linkie, Suyadi, P. Levang, and N. Leader-Williams. 2009b. Three decades of deforestation in southwest Sumatra: Effects of coffee prices, law enforcement and rural poverty. Biological Conservation **142**:597-605.

Geist, H. J., and E. F. Lambin. 2002. Proximate causes and underlying driving forces of tropical deforestation. Bioscience **52**:143-150.

Gibbs, H. K., A. S. Ruesch, F. Achard, M. K. Clayton, P. Holmgren, N. Ramankutty, and J. A. Foley. 2010. Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. Proceedings of the National Academy of Sciences of the United States of America **107**:16732-16737.

Gibson, L., T. M. Lee, L. P. Koh, B. W. Brook, T. A. Gardner, J. Barlow, C. A. Peres, C. J. A. Bradshaw, W. F. Laurance, T. E. Lovejoy, and N. S. Sodhi. 2011. Primary forests are irreplaceable for sustaining tropical biodiversity. Nature **478**:378-381.

Gomes, E. G., C. R. Grego, J. C. C. B. S. d. Mello, G. S. Valladares, J. A. d. C. Mangabeira, and E. E. d. Miranda. 2009. Dependência espacial da eficiência do uso da terra em assentamento rural na Amazônia. Produção **19**:417-432.

Grainger, A. 2008. Difficulties in tracking the long-term global trend in tropical forest area. Proceedings of the National Academy of Sciences **105**:818-823.

Grimm, V., U. Berger, F. Bastiansen, S. Eliassen, V. Ginot, J. Giske, J. Goss-Custard, T. Grand, S. K. Heinz, G. Huse, A. Huth, J. U. Jepsen, C. Jørgensen, W. M. Mooij, B. Müller, G. Pe'er, C. Piou, S. F. Railsback, A. M. Robbins, M. M. Robbins, E. Rossmanith, N. Rüger, E. Strand, S. Souissi, R. A. Stillman, R. Vabø, U. Visser, and D. L. DeAngelis. 2006. A standard protocol for describing individual-based and agent-based models. Ecological Modelling **198**:115-126.

Hall, C. A. S., H. Tian, Y. Qi, G. Pontius, and J. Cornell. 1995. Modelling spatial and temporal patterns of tropical land use change. Journal of Biogeography **22**:753-757.

Hansen, M. C., S. V. Stehman, and P. V. Potapov. 2010. Quantification of global gross forest cover loss. Proceedings of the National Academy of Sciences of the United States of America **107**:8650-8655.

Harris, N. L., S. Brown, S. C. Hagen, S. S. Saatchi, S. Petrova, W. Salas, M. C. Hansen, P. V. Potapov, and A. Lotsch. 2012a. Baseline Map of Carbon Emissions from Deforestation in Tropical Regions. Science **336**:1573-1576.

Harris, N. L., S. Brown, C. Stephen, A. Baccini, and R. A. Houghton. 2012b. Progress toward a consensus on carbon emissions from tropical deforestation. Winrock International and Woods Hole Research Center Washington, USA.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. Model Assessment and Selection: The Elements of Statistical Learning. Pages 219-259. Springer New York.

Hayashi, S., C. Souza Jr, and A. Verissimo. 2011. Mato Grosso SAD alerta. Available from http://www.imazon.org.br/publicacoes/transparencia-florestal/sad-alerta/sad-alerta-mato-grosso-abril-2011 (Accessed June 2011), Belém, Pará, Brazil.

Holly, K. G., B. Sandra, O. N. John, and A. F. Jonathan. 2007. Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environmental Research Letters **2**:045023.

Houghton, R. A. 2003. Why are estimates of the terrestrial carbon balance so different? Global Change Biology **9**:500-509.

Houghton, R. A. 2005. Tropical deforestation as a source of greenhouse gas emissions in P. Moutinho, and S. Schwartzman, editors. Tropical deforestation and climate change. Amazon Institute for Environmental Research.

Houghton, R. A., and J. L. Hackler. 1999. Emissions of carbon from forestry and land-use change in tropical Asia. Global Change Biology **5**:481-492.

Houghton, R. A., J. L. Hackler, and K. T. Lawrence. 1999. The U.S. Carbon Budget: Contributions from Land-Use Change. Science **285**:574-578.

Houghton, R. A., D. L. Skole, C. A. Nobre, J. L. Hackler, K. T. Lawrence, and W. H. Chomentowski. 2000. Annual fluxes of carbon from deforestation and regrowth in the Brazilian Amazon. Nature **403**:301-304.

Houghton, R. A., G. R. van der Werf, R. S. DeFries, M. C. Hansen, J. I. House, C. Le Quéré, J. Pongratz, and N. Ramankutty. 2012. Chapter G2 Carbon emissions from land use and land-cover change. Biogeosciences Discuss. **9**:835-878.

Huang, J., R. G. Pontius Jr, Q. Li, and Y. Zhang. 2012. Use of intensity analysis to link patterns with processes of land change from 1986 to 2007 in a coastal watershed of southeast China. Applied Geography **34**:371-384.

IBGE 2006. Censo Agropecuario 2006. Instituto Brasileiro de Geografia e Estatística, Rio de Janeiro, Brasil.

INPE. 2012a. Desflorestamento nos Municípios da Amazônia Legal. . Available from http://www.dpi.inpe.br/prodesdigital/municipios.html. (Accessed November 2010), São José dos Campos, São Paulo, Brazil.

INPE. 2012b. PRODES project - satellite monitoring of the Brazilian Amazon.

INPE. 2012c. Taxas anuais do desmatamento - 1988 até 2011.

IPAM. 2012. Plano de Ação para Prevenção e Controle do Desmatamento na Amazônia Legal – PPCDAM.

Jiang, Z. 2007. The road extension model in the Land change modeler for ecological sustainability of IDRISI. Proceedings of the 15th international symposium on advances in geographic information systems.

Joppa, L. N., S. R. Loarie, and S. L. Pimm. 2008. On the protection of “protected areas”. Proceedings of the National Academy of Sciences **105**:6673-6678.

Kaimowitz, D., B. Mertens, S. Wunder, and P. Pacheco. 2004. Hamburger connection fuels Amazon destruction. Centre for International Forestry Research.

Keith, H., B. G. Mackey, and D. B. Lindenmayer. 2009. Re-evaluation of forest biomass carbon stocks and lessons from the world's most carbon-dense forests. Proceedings of the National Academy of Sciences **106**:11635-11640.

Killeen, T. J., V. Calderon, L. Soria, B. Quezada, M. K. Steininger, G. Harper, L. A. Solo´rzano, and C. J. Tucker. 2007. Thirty Years of Land-cover Change in Bolivia. Ambio **36**:600 - 606.

Kinnaird, M. F., E. W. Sanderson, T. G. O'Brien, H. T. Wibisono, and G. Woolmer. 2003. Deforestation Trends in a Tropical Landscape and Implications for Endangered Large Mammals

Tendencias de Deforestación en un Paisaje Tropical y Sus Implicancias para Mamíferos Grandes en Peligro. Conservation Biology **17**:245-257.

Labarta, R. A., D. S. White, and S. M. Swinton. 2008. Does Charcoal Production Slow Agricultural Expansion into the Peruvian Amazon Rainforest? World Development **36**:527-540.

Labrecque, S., R. Fournier, J. Luther, and D. Piercey. 2006. A comparison of four methods to map biomass from Landsat-TM and inventory data in western Newfoundland. Forest Ecology and Management **226**:129-144.

Lambin, E. F. 1997. Modelling and monitoring land-cover change processes in tropical regions. Progress in Physical Geography **21**:375-393.

Lambin, E. F., H. J. Geist, and E. Lepers. 2003. DYNAMICS OF LAND-USE AND LAND-COVER CHANGE IN TROPICAL REGIONS. Annu. Rev. Environ. Resour. **28**:205 - 241.

Lambin, E. F., B. L. Turner, H. J. Geist, S. B. Agbola, A. Angelsen, J. W. Bruce, O. T. Coomes, R. Dirzo, G. n. Fischer, C. Folke, P. S. George, K. Homewood, J. Imbernon, R. Leemans, X. Li, E. F. Moran, M. Mortimore, P. S. Ramakrishnan, J. F. Richards, H. Sk√•nes, W. Steffen, G. D. Stone, U. Svedin, T. A. Veldkamp, C. Vogel, and J. Xu. 2001. The causes of land-use and land-cover change: moving beyond the myths. Global Environmental Change **11**:261-269.

Lapola, D. M., R. Schaldach, J. Alcamo, A. Bondeau, J. Koch, C. Koelking, and J. A. Priess. 2010. Indirect land-use changes can overcome carbon savings from biofuels in Brazil.

Proceedings of the National Academy of Sciences of the United States of America **107**:3388-3393.

Lapola, D. M., R. Schaldach, J. Alcamo, A. Bondeau, S. Msangi, J. A. Priess, R. Silvestrini, and B. S. Soares-Filho. 2011. Impacts of Climate Change and the End of Deforestation on Land Use in the Brazilian Legal Amazon. Earth Interactions **15**:1-29.

Laporte, N. T., J. A. Stabach, R. Grosch, T. S. Lin, and S. J. Goetz. 2007. Expansion of Industrial Logging in Central Africa. Science **316**:1451-1451.

Laurance, W. F. 1999. Reflections on the tropical deforestation crisis. Biological Conservation **91**:109-117.

Laurance, W. F. 2004. Forest-climate interactions in fragmented tropical landscapes. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **359**:345-352.

Laurance, W. F. 2007. Switch to Corn Promotes Amazon Deforestation. Science **318**:1721-1721.

Laurance, W. F., A. K. M. Albernaz, G. Schroth, P. M. Fearnside, S. Bergen, E. M. Venticinque, and C. Da Costa. 2002. Predictors of deforestation in the Brazilian Amazon. Journal of Biogeography **29**:737-748.

Laurance, W. F., M. A. Cochrane, S. Bergen, P. M. Fearnside, P. Delamonica, C. Barber, S. D'Angelo, and T. Fernandes. 2001a. Environment - The future of the Brazilian Amazon. Science **291**:438-439.

Laurance, W. F., M. A. Cochrane, S. Bergen, P. M. Fearnside, P. Delamonica, C. Barber, S. D'Angelo, and T. Fernandes. 2001b. The future of the Brazilian Amazon. Science **291**:438-439.

Le Quere, C., M. R. Raupach, J. G. Canadell, G. Marland, and et al. 2009. Trends in the sources and sinks of carbon dioxide. Nature Geosci **2**:831-836.

Lewis, D. J., A. J. Plantinga, and J. Wu. 2009a. Targeting Incentives to Reduce Habitat Fragmentation. American Journal of Agricultural Economics **91**:1080-1096.

Lewis, S. L., G. Lopez-Gonzalez, B. Sonke, K. Affum-Baffoe, T. R. Baker, L. O. Ojo, O. L. Phillips, J. M. Reitsma, L. White, J. A. Comiskey, M.-N. D. K, C. E. N. Ewango, T. R. Feldpausch, A. C. Hamilton, M. Gloor, T. Hart, A. Hladik, J. Lloyd, J. C. Lovett, J.-R. Makana, Y. Malhi, F. M. Mbago, H. J. Ndangalasi, J. Peacock, K. S. H. Peh, D. Sheil, T. Sunderland, M. D. Swaine, J. Taplin, D. Taylor, S. C. Thomas, R. Votere, and H. Woll. 2009b. Increasing carbon storage in intact African tropical forests. Nature **457**:1003-1006.

Liu, H., J. Shi, and E. Erdem. 2012. An Integrated Wind Power Forecasting Methodology: Interval Estimation Of Wind Speed, Operation Probability Of Wind Turbine, And Conditional Expected Wind Power Output Of A Wind Farm. International Journal of Green Energy **10**:151-176.

Loarie, S. R., G. P. Asner, and C. B. Field. 2009. Boosted carbon emissions from Amazon deforestation. Geophys. Res. Lett. **36**:L14810.

Lopez, S., and R. Sierra. 2010. Agricultural change in the Pastaza River Basin: A spatially explicit model of native Amazonian cultivation. Applied Geography **30**:355-369.

Lorena, R. B., and E. F. Lambin. 2009. The spatial dynamics of deforestation and agent use in the Amazon. Applied Geography **29**:171-181.

Ludeke, A. K., R. C. Maggio, and L. M. Reid. 1990. An Analysis of Anthropogenic Deforestation Using Logistic-Regression and Gis. Journal of Environmental Management **31**:247-259.

Maeda, E. E., C. M. de Almeida, A. de Carvalho Ximenes, A. R. Formaggio, Y. E. Shimabukuro, and P. Pellikka. 2011. Dynamic modeling of forest conversion: Simulation of past and future scenarios of rural activities expansion in the fringes of the Xingu National Park, Brazilian Amazon. International Journal of Applied Earth Observation and Geoinformation **13**:435-446.

Malhi, Y., and J. Grace. 2000. Tropical forests and atmospheric carbon dioxide. Trends in Ecology & Evolution **15**:332-337.

Malhi, Y., P. Meir, and S. Brown. 2002. Forests, carbon and global climate. Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **360**:1567-1591.

Malingreau, J., H. Eva, and E. de Miranda. 2012. Brazilian Amazon: A Significant Five Year Drop in Deforestation Rates but Figures are on the Rise Again. AMBIO: A Journal of the Human Environment **41**:309-314.

Mangabeira, J. A. C., E. E. Miranda, A. Dorado, and M. Guimaraes. 1998. Dinâmica do desmatamento em projetos de colonização: o caso de Machadinho D´Oeste em Rondônia. Page 7. Pesquisa em Andamento. EMBRAPA, Campinas, Brazil.

Mann, M. L., R. K. Kaufmann, D. Bauer, S. Gopal, M. D. Vera-Diaz, D. Nepstad, F. Merry, J. Kallay, and G. S. Amacher. 2010. The economics of cropland conversion in Amazonia: The importance of agricultural rent. Ecological Economics **69**:1503-1509.

McGuffie, K., A. Henderson-Sellers, H. Zhang, T. B. Durbidge, and A. J. Pitman. 1995. Global climate sensitivity to tropical deforestation. Global and Planetary Change **10**:97-128.

Mello, R., and P. Hildebrand. 2012. Modeling Effects of Climate Change Policies on Small Farmer Households in the Amazon Basin, Brazil. Journal of Sustainable Forestry **31**:59-79.

Mena, C. F., S. J. Walsh, B. G. Frizzelle, Y. Xiaozheng, and G. P. Malanson. 2011. Land use change on household farms in the Ecuadorian Amazon: Design and implementation of an agent-based model. Applied Geography **31**:210-222.

Mertens, B., and E. F. Lambin. 1997. Spatial modelling of deforestation in southern Cameroon : Spatial disaggregation of diverse deforestation processes. Applied Geography **17**:143-162.

Messina, J. P., T. P. Evans, S. M. Manson, A. M. Shortridge, P. J. Deadman, and P. H. Verburg. 2008. Complex systems models and the management of error and uncertainty. Journal of Land Use Science **3**:11-25.

Messina, J. P., and S. J. Walsh. 2001. 2.5D Morphogenesis: modeling landuse and landcover dynamics in the Ecuadorian Amazon. Plant Ecology **156**:75-88.

Metzger, J. P. 2002. Landscape dynamics and equilibrium in areas of slash-and-burn agriculture with short and long fallow period (Bragantina region, NE Brazilian Amazon). Landscape Ecology **17**:419-431.

Michalski, F., J. P. Metzger, and C. A. Peres. 2010. Rural property size drives patterns of upland and riparian forest retention in a tropical deforestation frontier. Global Environmental Change **20**:705-712.

Michalski, F., C. A. Peres, and I. R. Lake. 2008. Deforestation dynamics in a fragmented region of southern Amazonia: evaluation and future scenarios. Environmental Conservation **35**:93-103.

Miettinen, J., C. Shi, and S. C. Liew. 2011. Deforestation rates in insular Southeast Asia between 2000 and 2010. Global Change Biology **17**:2261-2270.

Millington, J. A., G. W. Perry, and R. Romero-Calcerrada. 2007. Regression Techniques for Examining Land Use/Cover Change: A Case Study of a Mediterranean Landscape. Ecosystems **10**:562-578.

Minnen, J., K. Klein Goldewijk, E. Stehfest, B. Eickhout, G. Drecht, and R. Leemans. 2009. The importance of three centuries of land-use change for the global and regional terrestrial carbon cycle. Climatic Change **97**:123-144.

Miranda, E. E. 2012. Sustentabilidade Agrícola na Amazonia - 23 anos de monitoramento da agricultura em Machadinho d'Oeste (RO). EMBRAPA.

Miranda, E. E., J. A. C. Mangabeira, and C. R. Grego. 2008. Desmistificação da terra: o perfil em 2005 dos produtores rurais de Machadinho d´Oeste (R0), 23 anos depois. Page 65. EMBRAPA, Campinas, Brazil.

Miranda, E. E., J. A. C. Mangabeira, V. A. Pires, and J. P. Franzin. 1999. Evolução agro-sócio-econômica dos agricultores e da agricultura de Machadinho d'Oeste em Rondônia, entre 1986 e 1999. Page 6. Pesquisa em Andamento. EMBRAPA, Campinas, Brazil.

Moran, E. 1993. Deforestation and land use in the Brazilian Amazon. Human Ecology **21**:1-21.

Moreira, E., S. Costa, A. P. Aguiar, G. Camara, and T. Carneiro. 2009. Dynamical coupling of multiscale land change models. Landscape Ecology **24**:1183-1194.

Morton, D. C., R. S. DeFries, Y. E. Shimabukuro, L. O. Anderson, E. Arai, F. D. Espirito-Santo, R. Freitas, and J. Morisette. 2006. Cropland expansion changes deforestation dynamics in the southern Brazilian Amazon. Proceedings of the National Academy of Sciences of the United States of America **103**:14637-14641.

Müller, R., D. Müller, F. Schierhorn, and G. Gerold. 2011. Spatiotemporal modeling of the expansion of mechanized agriculture in the Bolivian lowland forests. Applied Geography **31**:631-640.

Murray, J. D. 2002. Mathematical Biology: I. An Introduction, Third Edition. Springer.

Myers, N. 1988. Threatened Biotas: 'Hotspots' in Tropical Forests. The Environmentalist **8**:187-208.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, and J. Kent. 2000. Biodiversity hotspots for conservation priorities. Nature **403**:853-858.

Nelson, E., H. Sander, P. Hawthorne, M. Conte, D. Ennaanay, S. Wolny, S. Manson, and S. Polasky. 2010. Projecting Global Land-Use Change and Its Effect on Ecosystem Service Provision and Biodiversity with Simple Models. PLoS ONE **5**:e14327.

Nepstad, D., G. Carvalho, A. Cristina Barros, A. Alencar, J. Paulo Capobianco, J. Bishop, P. Moutinho, P. Lefebvre, U. Lopes Silva Jr, and E. Prins. 2001. Road paving, fire regime feedbacks, and the future of Amazon forests. Forest Ecology and Management **154**:395-407.

Nepstad, D., A. Moreira, A. Veríssimo, P. Lefebvre, P. Schlesinger, C. Potter, C. Nobre, A. Setzer, T. Krug, A. C. Barros, A. Alencar, and J. R. Pereira. 1998. Forest Fire Prediction and Prevention in the Brazilian Amazon. Conservation Biology **12**:951-953.

Nepstad, D., S. Schwartzman, B. Bamberger, M. Santilli, D. Ray, P. Schlesinger, P. Lefebvre, A. Alencar, E. Prinz, G. Fiske, and A. Rolla. 2006a. Inhibition of Amazon deforestation and fire by parks and indigenous lands. Conservation Biology **20**:65-73.

Nepstad, D., B. S. Soares, F. Merry, A. Lima, P. Moutinho, J. Carter, M. Bowman, A. Cattaneo, H. Rodrigues, S. Schwartzman, D. G. McGrath, C. M. Stickler, R. Lubowski, P. Piris-Cabezas, S. Rivero, A. Alencar, O. Almeida, and O. Stella. 2009. The End of Deforestation in the Brazilian Amazon. Science **326**:1350-1351.

Nepstad, D. C., C. M. Stickler, and O. T. Almeida. 2006b. Globalization of the Amazon soy and beef industries: Opportunities for conservation. Conservation Biology **20**:1595-1603.

Nobre, C. A., P. J. Sellers, and J. Shukla. 1991. Amazonian deforestation and regional climate change. Journal of Climate **4**:957-988.

Pacheco, P. 2009. Agrarian Reform in the Brazilian Amazon: Its Implications for Land Distribution and Deforestation. World Development **37**:1337-1347.

Parker, D., D. Brown, J. G. Polhill, S. M. Manson, and P. J. Deadman. 2008. Illustrating a new 'conceptual design pattern' for agent-based models and land use via five case studies: the MR POTATOHEAD framework. Pages 29-62 in A. L. Paredes, and C. H. Iglesias, editors. Agent-based Modelleling in Natural Resource Management. Universidad de Valladolid, Valladolid, Spain.

Pereira, D., D. Santos, M. Vedoveto, J. Guimaraes, and A. Verissimo. 2010a. Factos florestais da Amazonia. Instituto do Homem e Meio Ambiente da Amazônia, Belém, Brazil.

Pereira, H. M., P. W. Leadley, V. Proença, R. Alkemade, J. P. W. Scharlemann, J. F. Fernandez-Manjarrés, M. B. Araújo, P. Balvanera, R. Biggs, W. W. L. Cheung, L. Chini, H. D. Cooper, E. L. Gilman, S. Guénette, G. C. Hurtt, H. P. Huntington, G. M. Mace, T. Oberdorff, C. Revenga, P. Rodrigues, R. J. Scholes, U. R. Sumaila, and M. Walpole. 2010b. Scenarios for Global Biodiversity in the 21st Century. Science.

Perz, S. G., M. M. Caldas, E. Arima, and R. J. Walker. 2007. Unofficial Road Building in the Amazon: Socioeconomic and Biophysical Explanations. Development and Change **38**:529-551.

Pfaff, A., J. Robalino, R. Walker, S. Aldrich, M. Caldas, E. Reis, S. Perz, C. Bohrer, E. Arima, W. Laurance, and K. Kirby. 2007. Roads and deforestation in the Brazilian Amazon. Journal of Regional Science **47**:109-123.

Pfaff, A. S. P. 1999. What drives deforestation in the Brazilian Amazon? Evidence from satellite and socioeconomic data. Journal of Environmental Economics and Management **37**:26-43.

Phillips, O. L., Y. Malhi, N. Higuchi, W. F. Laurance, P. V. Nunez, R. M. Vasquez, S. G. Laurance, L. V. Ferreira, M. Stern, S. Brown, and J. Grace. 1998. Changes in the carbon balance of tropical forests: Evidence from long-term plots. Science **282**:439-442.

Pijanowski, B. C., D. G. Brown, B. A. Shellito, and G. A. Manik. 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. Computers, Environment and Urban Systems **26**:553-575.

Polhill, J. G., and N. Gotts. 2009. Ontologies for transparent integrated human-natural system modelling. Landscape Ecology **24**:1255-1267.

Pontius Jr, R. G., D. Huffaker, and K. Denman. 2004a. Useful techniques of validation for spatially explicit land-change models. Ecological Modelling **179**:445-461.

Pontius Jr, R. G., E. Shusas, and M. McEachern. 2004b. Detecting important categorical land changes while accounting for persistence. Agric Ecosyst Environ **101**:18-18.

Pontius, R. G., and M. Millones. 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing **32**:4407-4429.

Pontius, R. G., R. Walker, R. Yao-Kumah, E. Arima, S. Aldrich, M. Caldas, and D. Vergara. 2007. Accuracy Assessment for a Simulation Model of Amazonian Deforestation. Annals of the Association of American Geographers **97**:677-695.

Prasetyo, L. B., C. I. Wijaya, and Y. Setiawan. 2011. Spatial Model Approach for Deforestation: Case Study in Java Island, Indonesia. Pages 376-387. Land Use, Climate Change and Biodiversity Modeling: Perspectives and Applications. IGI Global.

Prentice, I., G. Farquhar, M. Fasham, M. Goulden, M. Heimann, H. Kheshi, L. Quere, R. Scholes, D. Wallace, and D. Archer. 2001. The carbon cycle and atmospheric carbon dioxide.

Ramankutty, N., H. K. Gibbs, F. Achard, R. Defries, J. A. Foley, and R. A. Houghton. 2007. Challenges to estimating carbon emissions from tropical deforestation. Global Change Biology **13**:51-66.

Ramos da Silva, R., D. Werth, and R. Avissar. 2008. Regional Impacts of Future Land-Cover Changes on the Amazon Basin Wet-Season Climate. Journal of Climate **21**:1153-1170.

Ribeiro, B., A. Veríssimo, and K. Pereira. 2006. Deforestation in Protected Areas in the Brazilian Amazon: the case of Rondonia. State of the Amazon. Imazon, Belém, Brazil.

Rosa, I. M. D., S. E. Ahmed, and R. M. Ewers. under review-a. The transparency, reliability and utility of land-use and land-cover change models Global Change Biology.

Rosa, I. M. D., C. Souza Jr., and R. M. Ewers. 2012. Changes in size of deforested patches in the Brazilian Amazon Conservation Biology **26**:932-937.

Rosa, I. M. D., C. Souza Jr., D. Purves, and R. M. Ewers. under review-b. Predictive modelling of contagious deforestation in the Brazilian Amazon. PLOS One.

Rudel, T. K. 2005. Tropical forests. Regional paths of destruction and regeneration in the late twentieth century. Columbia University Press, United States of America.

Rudel, T. K. 2007. Changing agents of deforestation: From state-initiated to enterprise driven processes, 1970–2000. Land Use Policy **24**:35-41.

Rudel, T. K., R. Defries, G. P. Asner, and W. F. Laurance. 2009. Changing Drivers of Deforestation and New Opportunities for Conservation. Conservation Biology **23**:1396-1405.

Rudorff, B. F. T., M. Adami, D. A. Aguiar, M. A. Moreira, M. P. Mello, L. Fabiani, D. F. Amaral, and B. M. Pires. 2011. The Soy Moratorium in the Amazon Biome Monitored by Remote Sensing Images. Remote Sensing **3**:185-202.

Rykiel Jr., E. J. 1996. Testing ecological models: the meaning of validation. Ecological Modelling **90**:229 - 244.

Saatchi, S. S., N. L. Harris, S. Brown, M. Lefsky, E. T. A. Mitchard, W. Salas, B. R. Zutta, W. Buermann, S. L. Lewis, S. Hagen, S. Petrova, L. White, M. Silman, and A. Morel. 2011. Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences.

Salati, E., and P. B. Vose. 1984. Amazon Basin: A System in Equilibrium. Science **225**:129-138.

Sangermano, F., J. R. Eastman, and H. Zhu. 2010. Similarity Weighted Instance-based Learning for the Generation of Transition Potentials in Land Use Change Modeling. Transactions in GIS **14**:569-580.

Sangermano, F., J. Toledano, and J. R. Eastman. 2012. Land cover change in the Bolivian Amazon and its implications for REDD+ and endemic biodiversity. Landscape Ecology **27**:571-584.

Sarkar, S., K. Crews-Meyer, K. Young, C. Kelley, and A. Moffett. 2009. A dynamic graph automata approach to modeling landscape change in the Andes and the Amazon. Environment and Planning B: Planning and Design **36**:300-318.

Sasaki, N., W. Knorr, D. R. Foster, H. Etoh, H. Ninomiya, S. Chay, S. Kim, and S. Sun. 2009. Woody biomass and bioenergy potentials in Southeast Asia between 1990 and 2020. Applied Energy **86, Supplement 1**:S140-S150.

Silvestrini, R. A., B. S. Soares-Filho, D. Nepstad, M. Coe, H. Rodrigues, and R. Assunção. 2011. Simulating fire regimes in the Amazon in response to climate change and deforestation. Ecological Applications **21**:1573-1590.

Smith, J., P. van de Kop, K. Reategui, I. Lombardi, C. Sabogal, and A. Diaz. 1999. Dynamics of secondary forests in slash-and-burn farming: interactions among land use types in the Peruvian Amazon. Agriculture, Ecosystems & Environment **76**:85-98.

Smith, M. J., M. C. Vanderwel, V. Lyutsarev, S. Emmott, and D. W. Purves. 2012. The climate dependence of the terrestrial carbon cycle; including parameter and structural uncertainties. Biogeosciences Discuss. **9**:13439-13496.

Soares-Filho, B., A. Alencar, D. Nepstad, G. Cerqueira, M. D. V. Diaz, S. Rivero, L. Solorzano, and E. Voll. 2004. Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: the Santarem-Cuiaba corridor. Global Change Biology **10**:745-764.

Soares-Filho, B., P. Moutinho, D. Nepstad, A. Anderson, H. Rodrigues, R. Garcia, L. Dietzsch, F. Merry, M. Bowman, L. Hissa, R. Silvestrini, and C. Maretti. 2010. Role of Brazilian Amazon protected areas in climate change mitigation. Proceedings of the National Academy of Sciences of the United States of America **107**:10821-10826.

Soares-Filho, B. S., G. C. Cerqueira, and C. L. Pennachin. 2002. DINAMICA - a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. Ecological Modelling **154**:217-235.

Soares-Filho, B. S., D. C. Nepstad, L. M. Curran, G. C. Cerqueira, R. A. Garcia, C. A. Ramos, E. Voll, A. McDonald, P. Lefebvre, and P. Schlesinger. 2006. Modelling conservation in the Amazon basin. Nature **440**:520-523.

Soler, L. D., P. Verburg, A. Veldkamp, M. I. S. Escada, and G. Camara. 2007. Statistical analysis and feedback exploration of land use change determinants at local scale in the Brazilian Amazon. Igarss: 2007 Ieee International Geoscience and Remote Sensing Symposium, Vols 1-12:3462-3465.

Sombroek, W. 2001. Spatial and temporal patterns of Amazon rainfall - Consequences for the planning of agricultural occupation and the protection of primary forests. Ambio **30**:388-396.

Steininger, M. K., C. Tucker, J. R. G. Townshend, T. J. Killeen, A. Desch, V. Bell, and P. Ersts. 2001 Tropical deforestation in the Bolivian Amazon Environ. Conserv. **28**:127–134

Steininger, M. K., C. J. Tucker, J. R. G. Townshend, T. J. Killeen, A. Desch, V. Bell, and P. Ersts. 2001. Tropical deforestation in the Bolivian Amazon. Environmental Conservation **28**:127-134.

Sterling, S. M., A. Ducharne, and J. Polcher. 2012. The impact of global land-cover change on the terrestrial water cycle. Nature Clim. Change **advance online publication**.

Svensson, J. 1998. Investment, property rights and political instability: Theory and evidence. European Economic Review **42**:1317-1341.

Tebaldi, C., and R. Knutti. 2007. The use of the multi-model ensemble in probabilistic climate projections. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **365**:2053-2075.

Thomson, M. C., F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. P. Morse, and T. N. Palmer. 2006. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature **439**:576-579.

Vadez, V., V. Reyes-Garcia, T. Huanca, and W. R. Leonard. 2008. Cash cropping, farm technologies, and deforestation: what are the connections? A model with empirical data from the Bolivian Amazon. Human Organization **67**:384 - 395.

van der Werf, G. R., D. C. D. Morton, R. S., J. G. J. Olivier, P. S. Kasibhatla, R. B. Jackson, G. J. Collatz, and J. T. Randerson. 2009. CO2 emissions from forest loss. Nature Geoscience **2**:737-738.

Van Minnen, J. G., K. K. Goldewijk, E. Stehfest, B. Eickhout, G. van Drecht, and R. Leemans. 2009. The importance of three centuries of land-use change for the global and regional terrestrial carbon cycle. Climatic Change **97**:123-144.

Van Vuuren, D. P., O. E. Sala, and H. M. Pereira. 2006. The future of vascular plant diversity under four global scenarios. Ecology and Society **11**:25.

Verburg, P., and A. Veldkamp. 2004. Projecting land use transitions at forest fringes in the Philippines at two spatial scales. Landscape Ecology **19**:77-98.

Verburg, P. H., W. Soepboer, R. Limpiada, M. V. O. Espaldon, M. A. Sharifa, and A. Veldkamp. 2002 Modelling the spatial dynamics of regional land use: The CLUE-S model. Environmental Management **30**:391-405.

Verburg, P. H., and A. Veldkamp. 2005. Introduction to the Special Issue on Spatial modeling to explore land use dynamics. International Journal of Geographical Information Science **19**:99-102.

Versace, V. L., D. Ierodiaconou, F. Stagnitti, and A. J. Hamilton. 2008. Appraisal of random and systematic land cover transitions for regional water balance and revegetation strategies. Agriculture, Ecosystems & Environment **123**:328-336.

Vourlitis, G. L., N. P. Filho, M. M. S. Hayashi, J. d. S. Nogueira, F. T. Caseiro, and J. H. Campelo, Jr. 2002. Seasonal variations in the evapotranspiration of a transitional tropical forest of Mato Grosso, Brazil. Water Resour. Res. **38**:1094.

Walker, R. 2003. Evaluating the Performance of Spatially Explicit Models. Photogrammetric Engineering & Remote Sensing **69**:1271–1278.

Walker, R., S. A. Drzyzga, Y. L. Li, J. G. Qi, M. Caldas, E. Arima, and D. Vergara. 2004. A behavioral model of landscape change in the Amazon Basin: The colonist case. Ecological Applications **14**:S299-S312.

Walker, R., E. Moran, and L. Anselin. 2000. Deforestation and Cattle Ranching in the Brazilian Amazon: External Capital and Household Processes. World Development **28**:683-699.

Walsh, S. J., J. P. Messina, C. F. Mena, G. P. Malanson, and P. H. Page. 2008. Complexity theory, spatial simulation models, and land use dynamics in the Northern Ecuadorian Amazon. Geoforum **39**:867-878.

Wassenaar, T., P. Gerber, P. H. Verburg, M. Rosales, M. Ibrahim, and H. Steinfeld. 2007a. Projecting land use changes in the Neotropics: The geography of pasture expansion into forest. Global Environmental Change-Human and Policy Dimensions **17**:86-104.

Wassenaar, T., P. Gerber, P. H. Verburg, M. Rosales, M. Ibrahim, and H. Steinfeld. 2007b. Projecting land use changes in the Neotropics: The geography of pasture expansion into forest. Global Environmental Change **17**:86-104.

WCRP. 2012. WCRP CMIP3 Multi-Model Dataset Archive at PCMDI.

Wearn, O., D. Reuman, and R. Ewers. 2012. Extinction debt and windows of conservation opportunity in the Brazilian Amazon. Science **337**:228-232.

White, R., and G. Engelen. 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Computers, Environment and Urban Systems **24**:383-400.

Wilkie, D., and N. T. Laporte. 2001. Forest area and deforestation in Central Africa: Current knowledge and future directions. Pages 119-138 in W. Weber, A. Vedder, S. Morland, and L. White, editors. African Rainforest Ecology and Conservation. Yale University Press, New Haven, CT.

Wilson, E. O. 1988. Biodiversity. National Academy of Sciences, Washington, D.C.

Wu, S.-S., X. Qiu, E. L. Usery, and L. Wang. 2009. Using Geometrical, Textural, and Contextual Information of Land Parcels for Classification of Detailed Urban Land Use. Annals of the Association of American Geographers **99**:76-98.

Wu, T.-J., P. Chen, and Y. Yan. 2013. The weighted average information criterion for multivariate regression model selection. Signal Processing **93**:49-55.

Yanai, A. M., P. M. Fearnside, P. M. L. d. A. Graça, and E. M. Nogueira. 2012. Avoided deforestation in Brazilian Amazonia: Simulating the effect of the Juma Sustainable Development Reserve. Forest Ecology and Management **282**:78-91.

Zhang, Q., D. Devers, A. Desch, C. Justice, and J. Townshend. 2005. MAPPING TROPICAL DEFORESTATION IN CENTRAL AFRICA. Environmental Monitoring and Assessment **101**:69-83.

Zhao, S., C. Peng, H. Jiang, D. Tian, X. Lei, and X. Zhou. 2006. Land use change in Asia and the ecological consequences. Ecological Research **21**:890-896.

Zimmerman, B., C. A. Peres, J. R. Malcolm, and T. Turner. 2001. Conservation and development alliances with the Kayapo of south-eastern Amazonia, a tropical forest indigenous people. Environmental Conservation **28**:10-22.

Appendices

Appendix A – Supporting Information for Chapter 3

A.1) Methodological results

The methodological details extracted from the 35 papers shown in Appendix A1 are shown in Table A1. The table displays the model methodological criteria which were used for comparisons throughout the review. The citation for each model is given, with the name of the model and type of model. If the model is agent based and if so what the agent is recorded as is whether or not the model is dynamic or static, deterministic and if it uses differing scenarios. The spatio-temporal resolution of the studies, the inputs, outputs, calibration and validation methods were also recoded. In addition, the number of years the model was run for was also noted.

Table A1 – Information extracted from each of the 35 papers reviewed. Methodological criteria, which were used for comparisons, include model type, spatial and temporal scales, type of model calibration, model and outputs validation. n/a refers to non-applicable; * Cell size was maintained in the units stated in each paper; ** Visual comparisons between observed and predicted datasets were considered as "no validation"; *** Model assessment is mentioned in the paper but no information on how it was performed; **** Landscape metric usually include fractal dimension, contagion index and number of patches per land use; ***** Validation used is similar to the one performed in this study where we assed match, omission (under predicting) and commission (over predicting); × Authors assume the model was validated in another paper, avoiding to perform a validation on their model outputs themselves; + no future predictions apart from years calibrated.

Reference	Model name	Model type	Individual level	Dynamic or Static	Deterministic or Stochastic	Region	If Brazil, state	Cell size*
Carpentier et al. (2000)	FaleBem	Optimisation Agent-based	farm	dynamic	deterministic	Brazil	Acre	not spatial
Dale et al. (1994)	DELTA	Agent-based	farm & individual	dynamic	stochastic	Brazil	Rondônia	100 m
de Barros et al. (2005)	n/a	Other (Markov chain)	n/a	dynamic	deterministic	Brazil	Rondônia	unclear
de Koning et al. (1999b)	CLUE	Statistic	n/a	dynamic	deterministic	Ecuador	n/a	5 min
de Koninget al. (1999a)	CLUE	Statistic	n/a	dynamic	deterministic	Ecuador	n/a	5 min
de Souza Soler et al. (2007)	CLUE-S	Statistic	n/a	static	stochastic	Brazil	Rondônia	250 m
Deadman et al. (2004)	LUCITA	Agent-based	farm	static	stochastic	Brazil	Pará	1 ha

Etter et al. (2006b)	IDRISI	Statistic	n/a	static	deterministic	Colombia	n/a	1 ha
Etter et al. (2006a)	n/a	Statistic	n/a	static	stochastic	Colombia	n/a	2 km
Evans et al. (2001)	n/a	Agent-based	parcel	dynamic	stochastic	Brazil	Pará	unclear
Labarta et al. (2008)	n/a	Optimisation Agent-based	household	dynamic	deterministic	Peru	n/a	not spatial
Lapola et al. (2011)	LandSHIFT	Other (hierarchal)	n/a	dynamic	deterministic	Brazil	All	5 arc
Lapola et al. (2010)	LandSHIFT	Other (hierarchal)	n/a	dynamic	deterministic	Brazil	All	5 arc
Laurance et al. (2001)	n/a	Statistic	n/a	static	deterministic	Brazil	All	unclear
Lopez et al. (2010)	n/a	Agent-based	household	static	deterministic	Ecuador	n/a	20 m
Maeda et al. (2010)	DINAMICA	CA	n/a	dynamic	stochastic	Brazil	Mato Grosso	100 m
Mann et al. (2010)	n/a	Statistic	n/a	static	deterministic	Brazil	Mato Grosso	232 m
Mena et al. (2011)	n/a	Agent-based	farm	dynamic	stochastic	Ecuador	n/a	unclear
Messina & Walsh (2001)	n/a	CA	farm	dynamic	stochastic	Ecuador	n/a	30 m
Michalski et al. (2008)	IDRISI	Other (Neural Network)	n/a	dynamic	deterministic	Brazil	Mato Grosso	unclear
Moreira et al. (2009)	TerraME	Other (hierarchical) Agent-based	farm	dynamic	deterministic	Brazil	ALL Pará	5 km 1 km
Muller et al.	n/a	Statistic	n/a	dynamic	stochastic	Bolivia	n/a	200m

(2011)								
Nepstad et al. (2010)	DINAMICA	CA	n/a	dynamic	stochastic	Brazil	All	unclear
Sarkar et al. (2009)	n/a	CA Agent-based	n/a	dynamic	stochastic	Peru	n/a	30 m
Soares-Filho et al. (2002)	DINAMICA	CA	n/a	dynamic	stochastic	Brazil	Mato Grosso	100 m
Soares-Filho et al. (2004)	DINAMICA	CA	n/a	dynamic	stochastic	Brazil	N. Mato Grosso S. Pará	250 m
Soares-Filho et al. (2006)	DINAMICA	CA	n/a	dynamic	stochastic	Brazil	All	1 km
Vadez et al. (2008)	n/a	Agent-based	household	static	deterministic	Bolivia	n/a	not spatial
Walker et al. (2004)	n/a	Agent-based	farm	dynamic	stochastic	Brazil	Pará	20 m
Walsh et al. (2008)	n/a	CA Agent-based	farm	dynamic	stochastic	Ecuador	n/a	30 m
Wassenaar et al. (2007)	CLUE	Statistic	n/a	dynamic	deterministic	South America	n/a	5 km
Silvestrini et al. (2011)	DINAMICA	CA	n/a	dynamic	stochastic	All Amazon	n/a	2 km
Sangermano et al. (2012)	IDRISI	Other (Neural Network)	n/a	dynamic	deterministic	Bolivia	n/a	1 km
Mello & Hildebrand (2012)	n/a	Optimisation	farm	static	deterministic	Brazil	Pará	not spatial
Yanai et al. (2012)	DINAMICA	CA	n/a	dynamic	stochastic	Brazil	Amazonas	250 m

Continuation of Table A1

Reference	Inputs Data-types	Inputs
Carpentier et al. (2000)	bio-economic	price, cost & market data (e.g. milk quota, transportation costs), biophysical data (yield, nutrients demands, nutrient accumulation, etc.), data about farmers' initial conditions (e.g. household composition and income)
Dale et al. (1994)	geo socio-economic	lot size, soil, vegetation, market, transport systems (roads) and decision variables
de Barros et al. (2005)	bio geo	land use change trajectories, hydrology, roads, elevation, land use maps (1984-2002)
de Koning et al. (1999b)	geo socio-economic	soil texture, slope, soil fertility, altitude, precipitation, distance to markets, distance to roads, distance to rivers, total population, rural population, urban population, population living in poverty (rural/total population), population that is illiterate (rural/total population), population working in agriculture (rural/total population)
de Koninget al. (1999a)	geo socio-economic	soil characteristics, climate, slope, demography, income levels, occupation, distance to roads and markets
de Souza et al. (2007)	geo socio-economic	Land use map, slope, geomorphology, litology, soils, land suitability, precipitation, cost to urban areas, cost to mining areas, cost to saw mills, population density, income per capita, number of people per district, protected areas
Deadman et al. (2004)	geo socio-economic	household composition, household capital, soil quality, burn quality, roads, plots, land cover, household arrival
Etter et al. (2006b)	bio geo	soil fertility, rivers, roads, topography, settlements, neighbour map of forest and secondary vegetation
Etter et al. (2006a)	bio geo	slope, soil fertility, moisture availability, rain days, distance to towns, roads, rivers, protected areas, regional boundaries, forest cover
Evans et al. (2001)	bio socio-economic	demography, household economics, land-use decision making, labour allocation and institutions; biophysical parameters (soil fertility, topography and hydrograph)
Labarta et al. (2008)	geo economic	labour, capital, household food security, market prices, infrastructure, land cover
Lapola et al.	bio geo socio-	land use map, vegetation type, slope, national market attraction, distance to paved roads,

(2011)	economic	distance to crop land, distance to all roads, crop/grass yield, population, crop prices, potential yields
Lapola et al. (2010)	geo socio-economic	land use map, crop & livestock demands, potential crop yields, population density, socioeconomic projections
Laurance et al. (2001)	geographic	forest cover, logging, mining, highways and roads, navigable rivers, fires vulnerability, protected areas and existing and planned infrastructures projects
Lopez et al. (2010)	bio geo socio	population pressure, soil quality, slope, travel time, distance to service areas and accessibility infrastructures
Maeda et al. (2010)	geographic	soil type, transportation costs, distance to prior deforestation, distance to river, distance to roads, distance to urban centres, protected areas, distance to pasture & crops
Mann et al. (2010)	geo-economic	distance to roads, rent, soybean revenue & cost, risk, slope, land cover (deforestation, pasture, savannah), distance to agricultural area
Mena et al. (2011)	geo socio-economic	parcel ID, farm ID, land use classes, age of land use, slope angle, distance to main road, distance to farm dwelling, household composition, mortality rates, birth rates, cattle, cocoa, coffee prices, household income, decision making module, maintenance costs, migration rates (4 main modules & 20 sub modules)
Messina & Walsh (2001)	geo	land use and land use change trajectories, roads, rivers, topography, soil
Michalski et al. (2008)	geo socio	land cover maps, distance to roads, distance to existing disturbances, transition potential, human population, bovine herd size.
Moreira et al. (2009)	bio geo socio-economic	land-use map, 40 environmental, demography, agriculture structure, technology, market connectivity
Muller et al. (2011)	bio geo-economic	rainfall, soil fertility, slope, transportation costs, distance to prior deforestation,
Nepstad et al. (2010)	bio -economic	potential rents of soybeans, cattle and timber production, biomass/vegetation
Sarkar et al. (2009)	n/a	transition probabilities/user set rules
Soares-Filho et al. (2002)	bio geo	soil, vegetation, altitude, distance to rivers, distance to roads, urban attraction, transition probabilities, land cover map, dist to previous deforest and regrowth
Soares-Filho et al.	bio geo	landscape map, vegetation, soil, altitude, slope, protected areas, distance to main and

(2004)		secondary roads, distance to forest, previously deforested area
Soares-Filho et al. (2006)	bio geo	biophysical variables, infrastructure, topography, rivers, soils, climate, roads, towns, markets
Vadez et al. (2008)	geo economic	rice area, walking time, road access, income, market dependence
Walker et al. (2004)	geo	number of deforestation events, associated magnitudes, distance from highways, distance from lot front
Walsh et al. (2008)	geo socio-economic	transition probabilities, population density, accessibility to roads, terrain (slope, aspect, soil moisture), farm income
Wassenaar et al. (2007)	bio geo socio-economic	rainfall, dry season length, altitude, slope, geology, soil depth, soil drainage, soil fertility, protected areas, national parks, other park, population density, population growth, topography index, flat area index, landscape fragmentation, cost of access from road, cost of access to market, proximity to fire, population density, population growth
Silvestrini et al. (2011)	geo	distance to deforestation or *cerrado*, distance to roads, distance to forest, distance to thorns, elevation, protected areas; climate
Sangermano et al. (2012)	bio geo economic	soil, ph, temperature, precipitation seasonality, soil texture, land cover type, forest accessibility: distance to roads, agriculture attraction: cost distance to populated area, cost dist to largest city and distance to other deforested areas
Mello & Hildebrand (2012)	geo economic	land use, consumption requirements, cash, labour, and farm size
Yanai et al. (2012)	geo	land cover map, dist to rivers, altitude/slope, vegetation, soil, protected areas, distance to roads, transition rates

Continuation of Table A1

Reference	Output	Calibration Method	Validation Measure	Years validated	Time Steps
Carpentier et al. (2000)	land-use trajectories (no spatial output)	group of experts / linear programming	quantitative comparison between land use trends observed and predicted	2	25
Dale et al. (1994)	land use trajectories, maps and spatial statistics	weights are set by the user based on empirical data / regression	validated elsewhere[×]	n/a	40
de Barros et al. (2005)	land use map and trajectory	Markov transition probabilities / linear models	no validation	n/a	10
de Koning et al. (1999b)	land use map and trajectory	multiple regression	correlation between predicted and census data (backcast and not spatial)	1	20
de Koninget al. (1999a)	land cover map,	multiple regression	validated elsewhere[×]	n/a	20
de Souza et al. (2007)	land cover change probability map	expert knowledge / logistic regression	ROC	1	20
Deadman et al. (2004)	land use change rates	regression	trends compared to results from other study	3	30
Etter et al. (2006b)	land cover map and trajectory	logistic regression	ROC	1	0[+]
Etter et al. (2006a)	land cover map and trajectory	logistic regression / classification trees	ROC	1	6
Evans et al. (2001)	land cover map and change trajectory	user defined rules	amount of forest compared between observed and predicted	4	50
Labarta et al. (2008)	land cover trajectory	linear programming	no validation	n/a	10
Lapola et al.	maps and trajectory	analytic hierarchy process /	fuzzy similarity index	1	44

(2011)		multi-criteria suitability analysis	kappa classification		
Lapola et al. (2010)	land-use map and livestock density maps	analytic hierarchy process / multi-criteria suitability analysis	ROC	1	17
Laurance et al. (2001)	land cover map	extrapolation of past trends	no validation	n/a	20
Lopez et al. (2010)	agriculture change probability map	step wise logistic multiple regression	error matrix and ROC	1	0[+]
Maeda et al. (2010)	agricultural conversion map	weights of evidence	fuzzy similarity indices	1	10
Mann et al. (2010)	agriculture change probability map	logistic regression	no validation[**]	n/a	0[+]
Mena et al. (2011)	land use map and trajectory	regression (OLS)	no validation[***]	n/a	25
Messina & Walsh (2001)	land use and land cover	user-defined rules	correlation between observed and predicted	1	27
Michalski et al. (2008)	land cover trajectory	neural networks	kappa index agreement	2	12
Moreira et al. (2009)	land cover map	regression models	no validation[**]	n/a	28
Muller et al. (2011)	agricultural conversion risk map	logistic regression	pseudo R^2 and ROC	1	20
Nepstad et al. (2010)	deforestation rates	weights of evidence	validated elsewhere[×]	n/a	30
Sarkar et al. (2009)	transition probabilities	graphs and rules	no validation	n/a	15
Soares-Filho et al. (2002)	land cover map	logistic regression	landscape metrics[****]	2	8
Soares-Filho et al. (2004)	land cover map	weights of evidence	validated elsewhere[×]	n/a	30
Soares-Filho et al. (2006)	land cover map	weights of evidence	validated elsewhere[×]	n/a	50

Vadez et al. (2008)	Land cover trajectories	regression	no validation	n/a	20
Walker et al. (2004)	land cover map	regression	match, over and under *****	1	25
Walsh et al. (2008)	land-use map	past data	match, over and under *****	1	26
Wassenaar et al. (2007)	land-use change map	expert knowledge / stepwise logistic regression	ROC	1	10
Silvestrini et al. (2011)	fire risk map trajectories	weights of evidence and logistic	ROC and fuzzy similarity index	3	40
Sangermano et al. (2012)	land cover map	multi-layer neural network	AUC and Kappa indices	1	46
Mello & Hildebrand (2012)	land change trajectories	linear programming	quantitative comparison between land use trends observed and predicted	1	5
Yanai et al. (2012)	land cover map	weights of evidence	fuzzy similarity index	1	42

A.2) Spatial and temporal scales

Within the sample of 35 models, farm-scale models (Walker et al. 2004; Walsh et al. 2008; Lopez & Sierra 2010) generally used cell sizes of 20 – 30 m, whereas sub-regional models (Soares-Filho et al. 2002; Mann et al. 2010) used cell sizes of 100 m – 1 km cell size and country-level models used cell sizes larger than 1 km, up to a maximum of 25 km (Wassenaar et al. 2007a; Moreira et al. 2009). This is a general trend and inevitably there are exceptions, for example, Soares-Filho et al. (2006) modelled the entire Amazon using a relatively small 1 km^2 resolution.

A.3) Drivers of deforestation in the Amazon

Models can incorporate dynamic inputs in three main ways. First, they can be specified outside of the LULC change model and 'fed' directly into the model at specified time points. This was the approach taken by Soares-Filho et al. (2006), who established a schedule of dates at which planned road developments were likely to be completed. Second, dynamic inputs may be modelled in tandem to the main LULC change model. For example, Moreira et al. (2009) ran three models (at three different spatial scales) in tandem and linked inputs and outputs of the models to each other via temporal and spatial couplers. Mena et al. (2011) also used three linked models (agriculture, demography and migration) to predict LULC change, with each of the models generating outputs that can be used as dynamic inputs by the other modules. Third, and most commonly, dynamic inputs may be the output from the previous time step of the model itself. For example, any model that predicts spatially explicit patterns of LULC change will use a LULC map as an input and then alter that map through time as the model runs. At each time step, the LULC map that was predicted from the time step previous will be used as the basis for the prediction in the next time step. For example, Maeda et al. (2011) used distance to pasture as an input, and as new pasture was created through time, the model used the location of those the new pastures to recalculate the distances.

A.4) Modelling road expansion in the Amazon

The importance of roads as an input for LULC change modelling has been repeatedly demonstrated (Geist & Lambin 2002; Pfaff et al. 2007; Fearnside 2008), but remains the central challenge in predicting future patterns of deforestation (Barlow et al. 2011). The importance of roads as a model input may, however, vary across space and through time. For example, Maeda et al. (2011) pointed out that while proximity to roads plays an important role in driving agricultural expansion, the importance of this input may be suppressed in areas with well-established road networks. Roads are a dynamic, spatially explicit phenomenon but are often treated as static, which is problematic because roads are the single strongest predictor of spatial patterns of deforestation. Not taking into account the development of new roads in the long term reduces the predictive power and accuracy of deforestation models. Messina and Walsh (2001) used a 1990 road map and suggested that incorporating an updated road map for 2000 would improve model predictions by considering changes to road coverage and road surface type. They further stipulated that the use of annual road information in simulations would be useful to better their predictions of urban land developments.

There are three spatially explicit models of road network expansion in the Amazon (Soares-Filho et al. 2006; Jiang 2007; Arima et al. 2008b), all of which use least-cost paths to determine the route a new road might take, but only one of which has been formally tested and validated against real-world data (Arima et al. 2008b). This model attempts to recreate the road building decisions made by the logging industry (Arima et al. 2008b), and did so with reasonable success. The model predictions fitted 7.6% of the actual network exactly, 50% of the predicted roads fell within 700 m of the actual network, and nearly all predicted roads (90%) fell within 5 km of the actual network.

The first of these models is a road-constructor module within the DINAMICA land use change model (Soares-Filho et al. 2006), which simulates the expansion of a secondary road network based on land 'attractiveness' (topography and soil type, used to determine the endpoint of a new road). Existing road density and average rates of road growth per time step are also used to help determine the amount and general location of new roads (Soares-Filho et al. 2004). The second road model, IDRISI's road extension module (Jiang 2007), is based on similar principles as DINAMICA's but produces a hierarchal road network by allowing different spatial structures for primary, secondary and tertiary roads, and incorporates the cost of converting different LULC types into a road in the calculation of the least-cost path (*i.e.* the cost of converting a forest to road is different to that of converting a field to road). The final road building model was developed in two stages and attempted to recreate the road building decisions made by the logging industry (Arima et al. 2005b; Arima et al. 2008a). In the original model, Arima et al. (2005b) predicted both destination determinate (where road destination are selected and a road is built from the chosen destination to the existing road network) and destination indeterminate roads (where roads simply grow out from the existing network with no fixed destinations). In the Arima et al. (2008a) model only destination determinate roads were considered.

A.5) Model calibration

Regression techniques are a common way to calibrate models, with 19 models employing some type of regression of which linear and logistic were the most commonly used (de Koning et al. 1999a; Soares-Filho et al. 2002; Mann et al. 2010; Mena et al. 2011). Most models relied on multiple regression to simultaneously estimate the effect of multiple inputs, with these models typically being simplified by a stepwise procedure. Other calibration methods include machine learning techniques such as classification trees (Etter et al. 2006b), neural networks (Michalski et al. 2008) and analytical hierarchy processes (Lapola et al. 2010). A final class of calibration techniques rely on a Bayesian approach, such as that incorporated within the recently updated DINAMICA EGO software (Soares-Filho et al. 2004; Soares-Filho et al. 2006; Maeda et al. 2011). In some cases, statistical models were combined with expert knowledge to retain inputs known to be important despite a model simplification process removing them (Soler et al. 2007). Expert knowledge was also used in calibration where researchers were aware of changes that were not reflected in past data, such as impending government plans to introduce subsidies for a particular crop (Wassenaar et al. 2007a; Vadez et al. 2008). The rate of LULC change is often modelled by extrapolation from LULC change rates in the recent past (Ferraz et al. 2005; Wassenaar et al. 2007a; Maeda et al. 2011). Similarly, the location of LULC change can be based on extrapolating historical patterns into the future. For example, Laurance et al. (2001b) carried out analyses that revealed 30% of deforestation occurred within 10 km of roads and that highways have further reaching effects on deforestation than do smaller roads. This information was then used to predict future deforestation around planned highway and road developments.

A.6) The difficulties of model validation

The most frequently used method of model validation is the Receiver Operating Characteristic (ROC) curve (Wassenaar et al. 2007a; Lapola et al. 2010; Müller et al. 2011). The ROC curve compares the model output image with a reference LULC image and evaluates the proportion of 'hits' and 'false alarms' (Eastman et al. 2005), together generating the Area Under the Curve (AUC) value, a measurable and comparable statistic of the validity of the model. Kappa statistics can also be used to validate model predictions, (Michalski et al. 2008), reflecting the proportional agreement between observed and predicted changes after agreement by chance (or random agreement) has been removed.

Similarly, some authors computed error matrices to determine the percentage of correct classifications (Lopez & Sierra 2010), correlations between land use classes (Messina & Walsh 2001), or to compare observed land-use trends quantitatively measured and those predicted by the model (Deadman et al. 2004). When models are based on a neighbourhood context, such as cellular automata models, comparisons on a pixel-by-pixel basis are infeasible (Maeda et al. 2011), so most authors chose to use a series of landscape metrics or fuzzy similarity measures to compare their model outputs with the observed data, considering indices such as contagion (Dale et al. 1994; Frohn et al. 1996; Soares-Filho et al. 2002), dominance (Dale et al. 1994), fractal dimension (Dale et al. 1994; Frohn et al. 1996; Soares-Filho et al. 2002), number of patches of each LULC class (Soares-Filho et al. 2002) and percentage of forest cleared (Frohn et al. 1996).

A.7) Quantitative assessment of model performance

Of the dataset we were able to compile, we had three sets of model predictions that covered the whole Amazon (Soares-Filho et al. 2006; Wassenaar et al. 2007a; Lapola et al. 2010), three for Brazil (Soler et al. 2007; Maeda et al. 2011; Yanai et al. 2012), two for Ecuador (Messina & Walsh 2001; Walsh et al. 2008), two for Bolivia (Vadez et al. 2008; Müller et al. 2011) and one for Colombia (Etter et al. 2006b). The results obtained from our quantitative model performance of the Soares-Filho *et al.* (2006), Wassenaar *et al.* (2007) and Yanai *et al.* (2012) model predictions are shown on Table A2.

Table A2 – Results from pixel-by-pixel and distance-based comparisons between observed deforestation from PRODES and predictions made by Soares-Filho et al. (2006), from both governance (GOV) and business-as-usual (BAU) scenarios, Wassenaar *et al*. (2007), and made by Yanai *et al.* (2012) for the baseline, with leakage and with reduced leakage scenarios, annually and cumulatively.

Model	**Validation Time Period**	**Pixel-by-pixel**			**Distance-based metric**			
		Match	**Omission**	**Commission**	**1px**	**5px**	**10px**	**50px**
Wassennar et al. (2007)	2000-2010	0.08	0.62	0.30	0.32	0.44	0.81	0.99
Yanai et al. (2012) baseline	2008-2009	0.02	0.76	0.22	0.03	0.21	0.43	0.80
	2009-2010	0.01	0.78	0.20	0.04	0.23	0.48	0.90
	2008-2010	0.04	0.77	0.19	0.09	0.35	0.55	0.86
Yanai et al. (2012) with leakage	2008-2009	0.03	0.75	0.22	0.04	0.21	0.42	0.78
	2009-2010	0.02	0.79	0.20	0.06	0.26	0.51	0.92
	2008-2010	0.05	0.76	0.18	0.11	0.37	0.59	0.86
Yanai et al. (2012) with reduced leakage	2008-2009	0.03	0.77	0.20	0.06	0.26	0.47	0.78
	2009-2010	0.03	0.78	0.19	0.06	0.24	0.48	0.91
	2008-2010	0.05	0.77	0.17	0.11	0.38	0.59	0.86

(continuation of Table A2)

Model	Time Period	Pixel-by-pixel			Distance-based metric			
		Match	Omission	Commission	1px	5px	10px	50px
Soares-Filho et al. (2006) BAU	2002-2003	0.02	0.56	0.42	0.08	0.44	0.71	0.97
	2003-2004	0.01	0.53	0.46	0.08	0.45	0.72	0.98
	2004-2005	0.01	0.49	0.50	0.08	0.47	0.76	0.99
	2005-2006	0.01	0.30	0.70	0.08	0.48	0.76	0.99
	2006-2007	0.01	0.30	0.69	0.09	0.49	0.77	0.99
	2007-2008	0.01	0.32	0.67	0.09	0.53	0.80	0.99
	2008-2009	0.00	0.28	0.72	0.05	0.37	0.62	0.99
	2009-2010	0.01	0.18	0.82	0.09	0.55	0.80	0.97
	2002-2003	0.02	0.56	0.42	0.08	0.44	0.71	0.97
	2002-2004	0.03	0.54	0.43	0.14	0.59	0.80	0.98
	2002-2005	0.04	0.51	0.45	0.20	0.66	0.83	0.99
	2002-2006	0.05	0.46	0.49	0.26	0.70	0.86	0.99
	2002-2007	0.06	0.42	0.51	0.31	0.74	0.87	0.99
	2002-2008	0.08	0.40	0.53	0.36	0.76	0.89	0.99
	2002-2009	0.08	0.37	0.55	0.38	0.77	0.88	0.99
	2002-2010	0.09	0.34	0.57	0.42	0.79	0.89	0.99
Soares-Filho et al. (2006) GOV	2002-2003	0.01	0.58	0.41	0.07	0.42	0.65	0.96
	2003-2004	0.01	0.56	0.43	0.06	0.39	0.65	0.97
	2004-2005	0.01	0.53	0.45	0.06	0.39	0.65	0.97
	2005-2006	0.01	0.35	0.64	0.07	0.40	0.64	0.97
	2006-2007	0.01	0.37	0.63	0.05	0.36	0.62	0.95
	2007-2008	0.01	0.40	0.59	0.06	0.37	0.62	0.96
	2008-2009	0.00	0.37	0.63	0.03	0.26	0.50	0.97
	2009-2010	0.00	0.25	0.74	0.06	0.37	0.63	0.95
	2002-2003	0.01	0.58	0.41	0.07	0.42	0.65	0.96
	2002-2004	0.03	0.56	0.41	0.13	0.54	0.74	0.97
	2002-2005	0.04	0.55	0.41	0.18	0.59	0.77	0.98
	2002-2006	0.05	0.50	0.45	0.22	0.63	0.79	0.98
	2002-2007	0.06	0.48	0.47	0.25	0.65	0.80	0.98
	2002-2008	0.07	0.46	0.48	0.28	0.66	0.80	0.98
	2002-2009	0.07	0.44	0.49	0.30	0.66	0.80	0.99
	2002-2010	0.08	0.42	0.51	0.33	0.68	0.81	0.99

Further, to illustrate the process of comparing observed and predicted deforestation on a pixel by pixel basis and how we obtained the metrics of match, omission and commission, figure A1 shows the comparisons made between PRODES observed deforestation between 2000 and 2010 and predicted made by Wassenaar *et al.* (2007); figure A2 and A3 show the comparisons made between PRODES observed deforestation between 2002 and 2010 and predicted made by Soares-Filho *et al.* (2006) for both the BAU and GOV scenarios, respectively. We were unable to perform any comparisons between observed deforestation and predictions made in locations other than the Brazilian Amazon, due to a lack of readily available, reliable time series data set of observed deforestation. Further, of the four Brazilian Amazon datasets comparisons could only be made on two (Soares-Filho et al. 2006; Wassenaar et al. 2007a), because Lapola et al. (2010) and Maeda et al. (2011) only provided outputs for 2020 and 2015, respectively.

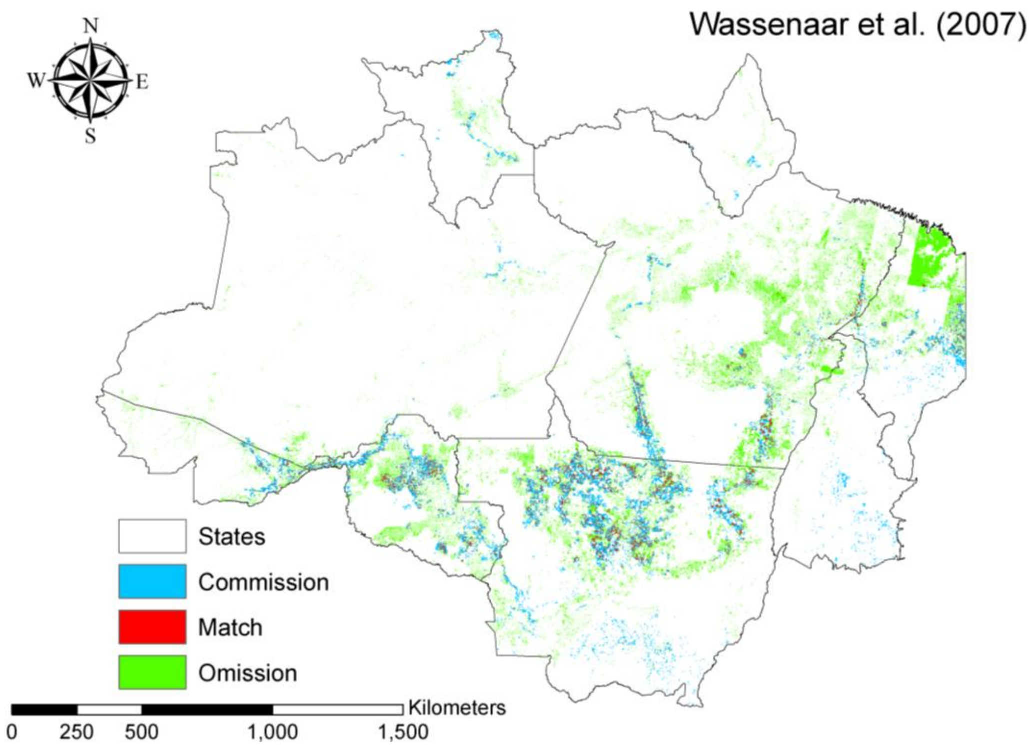


Figure A1 – Match, omission and commission between observed data from PRODES and model predictions made by Wassenaar et al. (2007a) from 2000 to 2010.

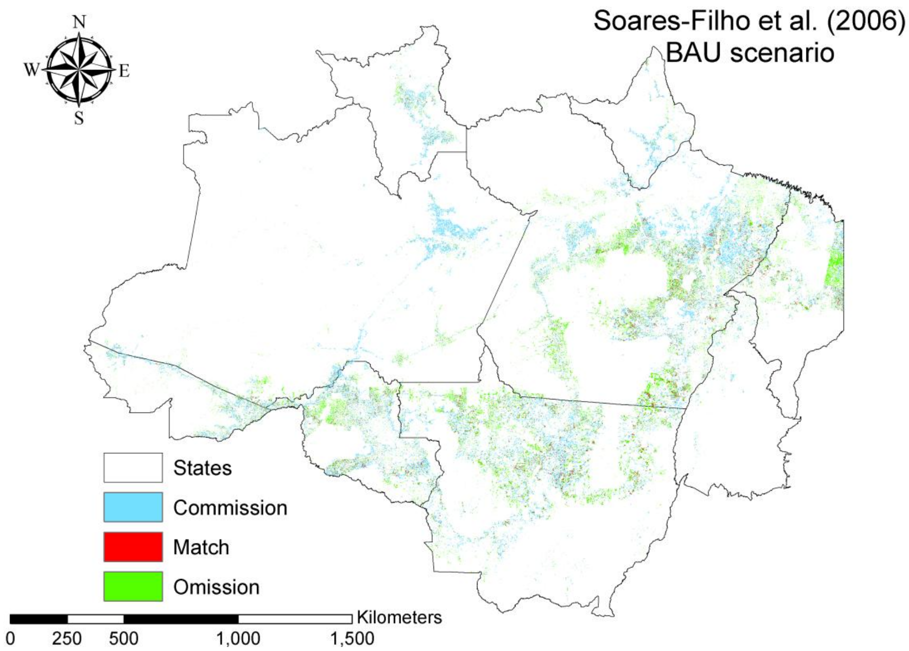


Figure A2 – Match, omission and commission between observed data from PRODES and model predictions by Soares-Filho et al. (2006) for the business-as-usual (BAU) scenario from 2002 to 2010.

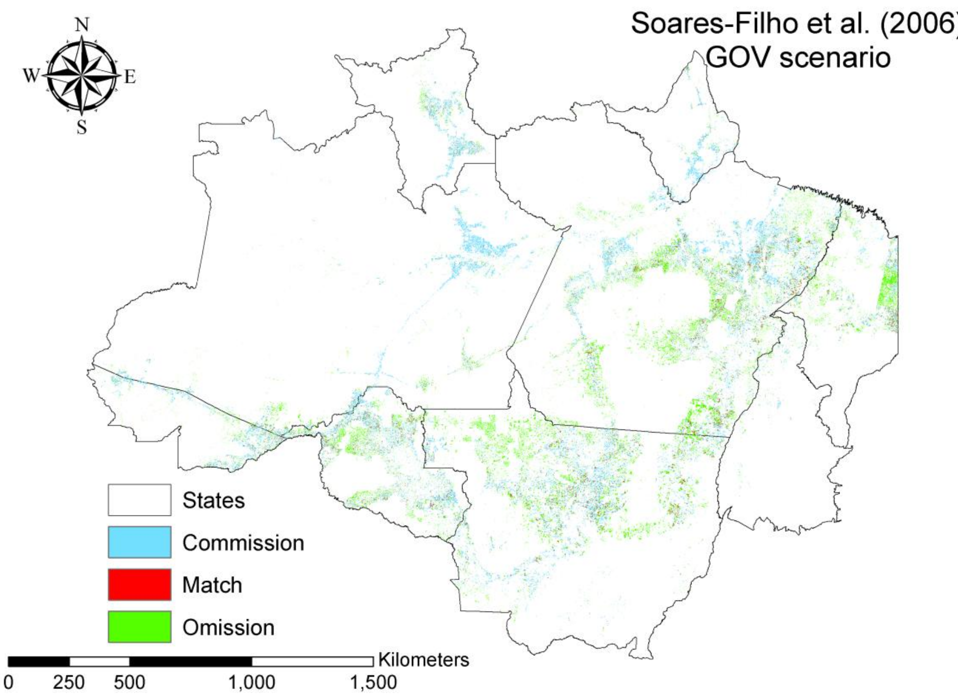


Figure A3 – Match, omission and commission between observed data PRODES and model predictions by Soares-Filho et al. (2006) for the governance (GOV) scenario from 2002 to 2010.

A.8) Appendix A1

1) Carpentier, C. L., S. A. Vosti, and J. Witcover. 2000. Intensified production systems on western Brazilian Amazon settlement farms: could they save the forest? Agriculture Ecosystems & Environment 82:73-88.

2) Dale, V. H., R. V. Oneill, F. Southworth, and M. Pedlowski. 1994. Modeling Effects of Land Management in the Brazilian Amazonian Settlement of Rondonia. Conservation Biology 8:196-206.

3) de Koning, G. H. J., A. Veldkamp, and L. O. Fresco. 1999a. Exploring changes in Ecuadorian land use for food production and their effects on natural resources. Journal of Environmental Management 57:221-237.

4) de Koning, G. H. J., P. H. Verburg, A. Veldkamp, and L. O. Fresco. 1999b. Multi-scale modelling of land use change dynamics in Ecuador. Agricultural Systems 61:77-93.

5) Deadman, P., D. Robinson, E. Moran, and E. Brondizio. 2004. Colonist household decisionmaking and land-use change in the Amazon Rainforest: an agent-based simulation. Environment and Planning B-Planning & Design 31:693-709.

6) Etter, A., C. McAlpine, K. Wilson, S. Phinn, and H. Possingham. 2006a. Regional patterns of agricultural land use and deforestation in Colombia. Agriculture Ecosystems & Environment 114:369-386.

7) Etter, A., C. McAlpine, S. Phinn, D. Pullar, and H. Possingham. 2006b. Unplanned land clearing of Colombian rainforests: Spreading like a disease? Landscape and urban planning 77:240-254.

8) Evans, T. P., A. Manire, F. de Castro, E. Brondizio, and S. McCracken. 2001. A dynamic model of household decision-making and parcel level landcover change in the eastern Amazon. Ecological Modelling 143:95-113.

9) Ferraz, S. F. D., C. A. Vettorazzi, D. M. Theobald, and M. V. R. Ballester. 2005. Landscape dynamics of Amazonian deforestation between 1984 and 2002 in central Rondonia, Brazil: assessment and future scenarios. Forest Ecology and Management 204:67-83.

10) Labarta, R. A., D. S. White, and S. M. Swinton. 2008. Does Charcoal Production Slow Agricultural Expansion into the Peruvian Amazon Rainforest? World Development 36:527-540.

11) Lapola, D. M., R. Schaldach, J. Alcamo, A. Bondeau, J. Koch, C. Koelking, and J. A. Priess. 2010. Indirect land-use changes can overcome carbon savings from biofuels in Brazil. Proceedings of the National Academy of Sciences of the United States of America 107:3388-3393.

12) Lapola, D.M., R. Schaldach, J. Alcamo, A. Bondeau, S. Msangi, J. A. Priess, R. Silvestrini, and B. Soares-Filho. 2011. Impacts of climate change and the end of deforestation on land use in the Brazilian Legal Amazon. Earth Interactions 15:1-29.

13) Laurance, W. F. 2001. The future of the Brazilian Amazon (vol 291, pg 438, 2001). Science 291:988-988.

14) Lopez, S., and R. Sierra. 2010. Agricultural change in the Pastaza River Basin: A spatially explicit model of native Amazonian cultivation. Applied Geography 30:355-369.

15) Maeda, E. E., C. M. de Almeida, A. de Carvalho Ximenes, A. R. Formaggio, Y. E. Shimabukuro, and P. Pellikka. 2011. Dynamic modeling of forest conversion: Simulation of past and future scenarios of rural activities expansion in the fringes of the Xingu National Park, Brazilian Amazon. International Journal of Applied Earth Observation and Geoinformation 13:435-446.

16) Mann, M. L., R. K. Kaufmann, D. Bauer, S. Gopal, M. D. Vera-Diaz, D. Nepstad, F. Merry, J. Kallay, and G. S. Amacher. 2010. The economics of cropland conversion in Amazonia: The importance of agricultural rent. Ecological Economics 69:1503-1509.

17) Mello, R., and Hildebrand, P. 2012. Modeling effects of climate change policies on small farmer households in the Amazon Basin, Brazil. Journal of Sustainable Forestry, 31: 59-79.

18) Mena, C. F., S. J. Walsh, B. G. Frizzelle, Y. Xiaozheng, and G. P. Malanson. 2011. Land use change on household farms in the Ecuadorian Amazon: Design and implementation of an agent-based model. Applied Geography 31:210-222.

19) Messina, J. P., and S. J. Walsh. 2001. 2.5D Morphogenesis: modeling landuse and landcover dynamics in the Ecuadorian Amazon. Plant Ecology 156:75-88.

20) Michalski, F., C. A. Peres, and I. R. Lake. 2008. Deforestation dynamics in a fragmented region of southern Amazonia: evaluation and future scenarios. Environmental Conservation 35:93-103.

21) Moreira, E., S. Costa, A. P. Aguiar, G. Camara, and T. Carneiro. 2009. Dynamical coupling of multiscale land change models. Landscape Ecology 24:1183-1194.

22) Müller, R., D. Müller, F. Schierhorn, and G. Gerold. 2011. Spatiotemporal modeling of the expansion of mechanized agriculture in the Bolivian lowland forests. Applied Geography 31:631-640.

23) Nepstad, D., B. S. Soares, F. Merry, A. Lima, P. Moutinho, J. Carter, M. Bowman, A. Cattaneo, H. Rodrigues, S. Schwartzman, D. G. McGrath, C. M. Stickler, R. Lubowski, P. Piris-Cabezas, S. Rivero, A. Alencar, O. Almeida, and O. Stella. 2009. The End of Deforestation in the Brazilian Amazon. Science 326:1350-1351.

24) Sangermano, F., Toledano, J., Eastman, J.R. 2012. Land cover change in the Bolivian Amazon and its implications for REDD+ and endemic biodiversity. Landscape Ecology 27: 571-584.

25) Sarkar, S., K. A. Crews-Meyer, K. R. Young, C. D. Kelley, and A. Moffett. 2009. A dynamic graph automata approach to modelling landscape change in the Andes and the Amazon. Environment and Planning B: Planning and Design 36: 301-318.

26) Silvestrini, R.A., Soares-Filho, B.S., Nepstad, D., Coe, M., Rodrigues, H.O., Assuncao, R. 2011. Simulating fire regimes in the Amazon in response to climate change and deforestation. Ecological Applications, 21:1573–1590.

27) Soares-Filho, B. S., G. C. Cerqueira, and C. L. Pennachin. 2002. DINAMICA - a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. Ecological Modelling 154:217-235.

28) Soares-Filho, B., A. Alencar, D. Nepstad, G. Cerqueira, M. D. V. Diaz, S. Rivero, L. Solorzano, and E. Voll. 2004. Simulating the response of land-cover changes to road paving and governance along a major Amazon highway: the Santarem-Cuiaba corridor. Global Change Biology 10:745-764.

29) Soares-Filho, B. S., D. C. Nepstad, L. M. Curran, G. C. Cerqueira, R. A. Garcia, C. A. Ramos, E. Voll, A. McDonald, P. Lefebvre, and P. Schlesinger. 2006. Modelling conservation in the Amazon basin. Nature 440:520-523.

30) Soler, L. D., P. Verburg, A. Veldkamp, M. I. S. Escada, and G. Camara. 2007. Statistical analysis and feedback exploration of land use change determinants at local scale in the Brazilian Amazon. 2007 IEEE International Geoscience and Remote Sensing Symposium, Vols 1-12:3462-3465

31) Vadez V, Reyes-Garcia V, Huanca T, Leonard WR (2008) Cash cropping, farm technologies, and deforestation: what are the connections? A model with empirical data from the Bolivian Amazon. *Human Organization,* **67,** 384 - 395.

32) Walker, R., S. A. Drzyzga, Y. L. Li, J. G. Qi, M. Caldas, E. Arima, and D. Vergara. 2004. A behavioral model of landscape change in the Amazon Basin: The colonist case. Ecological Applications 14:S299-S312.

33) Walsh, S. J., J. P. Messina, C. F. Mena, G. P. Malanson, and P. H. Page. 2008. Complexity theory, spatial simulation models, and land use dynamics in the Northern Ecuadorian Amazon. Geoforum 39:867-878.

34) Wassenaar, T., P. Gerber, P. H. Verburg, M. Rosales, M. Ibrahim, and H. Steinfeld. 2007. Projecting land use changes in the Neotropics: The geography of pasture expansion into forest. Global Environmental Change-Human and Policy Dimensions 17:86-104.

35) Yanai, A. M.; Feanside, P. M.; Graça, P. M. L. A; Nogueira, E. M. 2012. Avoided deforestation in Brazilian Amazonia: Simulating the effect of the Juma Sustainable Development Reserve. Forest Ecology and Management 282: 78-91.

A.9) Appendix A2

We identified five commercially or freely available modelling platforms that have been used to develop LULC models in the Amazon, which we summarise below:

1) The Land Change Modeller within IDRISI is a GIS modelling programme that makes predictions based on transition probabilities calculated from observed LULC change between two time points (Clark Labs 2007). Quantity predictions may either be made via a Markov chain analysis or by user defined transition probabilities that are derived from an external model, such as an econometric model (Clark Labs 2007). Prediction of locations more suitable for change can be made by extracting the relationships between historical changes and a set of driver variables through a Neural Networks, Logistic Regression or SimWeight (Sangermano et al. 2010) model. Sangermano et al. (2012), Michalski et al. (2008) as well as Etter et al. (2006a) make use of IDRISI in their predictions. Available on http://www.clarklabs.org/ (Accessed in March 2012).

2) DINAMICA EGO is an updated version of DINAMICA, is and utilises inbuilt simulation algorithms to perform LULC change modelling tasks. Coupled with VENSIM (system thinking software), alternative scenarios can be modelled in DINAMICA via a cellular automata modelling framework. DINAMICA was developed and used to predict LULC changes in the Amazon by Soares-Filho et al. (2002), Soares-Filho et al. (2004), Soares-Filho et al. (2006), Nepstad et al. (2009), Maeda et al. (2011), Silvestrini et al. (2011) and Yanai et al. (2012). Available on http://www.csr.ufmg.br/dinamica/ (Accessed in March 2012).

3) CLUE consists of two interlinked modules that estimate demand and then allocate that demand across space (Verburg et al. 2002). In the demand module, the total area needed for different land use types is calculated on the basis of national demands for various commodities, calculated through the sum of domestic consumption volumes and export volumes and under given constraints of population size, consumption patterns, import/export developments and agricultural productivity. The allocation module makes spatially-explicit predictions of land use change based upon the assumption that land use at each location is determined by a combination of bio-

geophysical characteristics, socio-economic and infrastructural factors (de Koning et al. 1999a; de Koning et al. 1999b). Wassenaar et al. (2007a) used a more updated version of CLUE, CLUE-S, to model land use changes across Central and South America. Soler et al. (2007) also used CLUE to predict deforestation in Rondônia. Available on http://www.ivm.vu.nl/en/Organisation/departments/spatial-analysis-decision-support/Clue/index.asp (Accessed in March 2012).

4) TerraME is a multi-scale modelling language that allows users to combine independent scale-specific models (with different methodologies, extents and resolutions) and run them at the same time. This hierarchical modelling platform admits bidirectional feedbacks by using both top-down and bottom-up linkages between multi-scaled models (Moreira et al. 2009). Available on http://www.terrame.org/doku.php (Accessed in March 2012).

5) Landshift is a spatially-explicit multi-scale land use change model (Lapola et al. 2010; Lapola et al. 2011). Its main aim is to use the interactions between economic, social and biophysical drivers to predict the evolution of a landscape under different scenarios and how that evolution impact human society. Available on http://www.usf.uni-kassel.de/cesr/index.php?option=com_project&task=view_detail&agid=27&lang=en (Accessed in February 2013).

Appendix B – Supporting information for Chapter 6

B.1) Data sources

The global land cover dataset used in this study is the Moderate Resolution Imaging Spectroradiometer (MODIS) product called "Land Cover Type Yearly L3 Global 500 m SIN Grid V051" (full information about this product can be found here: https://lpdaac.usgs.gov/products/modis_products_table/mcd12q1). It was downloaded from the National Aeronautics and Space Administration (NASA) Earth Observing data and information website (http://reverb.echo.nasa.gov/reverb/). This dataset includes ten annual global land cover maps (2001-2010), each with a legend of 16 classes (Friedl et al. 2002), and where tropical rainforests were identified by class number 2 ("Evergreen Broadleaf forest"). Only one transition is allowed in the model (deforested to forest). Additionally, we extracted the extent of agricultural lands (classes 12 and 14, respectively: "Croplands" and "Cropland/Natural vegetation mosaic") to be used as a proxy of what is already deforested. However, class 9 ("Savannas") clearly included all natural savannahs as well as deforested areas. Therefore, we removed these natural savannah areas by overlapping the Worldwide Wildlife Fund (WWF) dataset on world's biomes (http://www.worldwildlife.org/science/data/item6373.html). Class 9 pixels that were within the savannah biome defined by WWF were considered natural savannahs, otherwise they were considered deforested areas.

The data regarding global land cover change drivers was also collected from different sources. Global altitude and precipitation layers (30" grids) were downloaded from the WorldClim website (http://www.worldclim.org/current). Precipitation layer was used to calculate dry season length for each region based on Sombroek (2001) methodology. Road and river maps (vector data) for each country from the Digital Chart of the World were downloaded from DIVA-GIS website (http://www.diva-gis.org/gdata). Finally, a map of human population density in 2000 (30" grid) was obtained from the Gridded Population of the World website (http://sedac.ciesin.columbia.edu/gpw/) and protected areas (vector data) from the World Database of Protected Areas (http://protectedplanet.net/). The cell size used in this study was 0.01^{o} (~ 1 km^2) and the spatial reference was WGS84.

B.2) Modelling past deforestation

Our modelling approach is based on that of Rosa et al. (under review-b), who developed a dynamic and spatially-explicit model of deforestation to predict the potential magnitude and spatial pattern of deforestation. We constructed independent models that utilise a land cover map of time t with two land cover classes (forest, deforested), and for each pixel deforested at time *t*, the model calculates the probability of being forested at time *t-1*. This probability is defined as a logistic function:

$$P_{for,x,t} = 1/(1 + \exp[-\kappa_{x,t}]) \qquad (3)$$

such that as $\kappa_{x,t}$ goes from minus infinity to plus infinity, $P_{for,x,t}$ goes from 0 to 1.

The modelling process was initiated for each region with the same set of variables: land cover and proportion of forested neighbours in the annual transition being parameterised (both dynamic), distance to roads, distance to rivers, altitude, dry season length and population density (all static). For each of these models, given the input data described above, all possible models (29) at each of the nine annual transitions between 2002 through 2010 were fitted and a forward logistic stepwise regression was performed. Specifically, for every pixel that was deforested in the year of calibration (2002 through 2010) we determined the probability of being forest in the year before. Each of the steps of the stepwise regression differs only in the combination of variables included in the linear models for $\kappa_{x,t}$ which are a function of the driver variables affecting location *x* at time *t* (Rosa et al. under review-b).

To carry out the parameter estimation, all that was necessary was to define the log-likelihood:

$$\ell(X|s,\theta) = \sum_{x,t} log\{Z_{x,t}P_{for,x,t} + (1 - Z_{x,t})(1 - P_{for,x,t})\}3 \qquad (4)$$

where $Z_{x,t}$ is the observed 'forestation' at location *x* at time *t*, and *s* refers to one of the models considered. At each step of the forward stepwise regression, a cross-validation was carried out by parameterising the model against a randomly selected subset of 50% of locations and calculating the training likelihood, then calculating the test likelihood of the remaining 50% of the locations, using Eq. 4, for validation (Rosa et al. under review-b).

In order to do so, we re-applied Eq. 3 in each time step, recalculating the dynamics variables (*e.g.* fractional forest around each location *x*), and using a slightly different set of parameter values at each iteration (to incorporate parameter uncertainty) drawn using a Gaussian distribution from the credible interval determined above. This provided an updated $P_{for,x,t}$ for each location *x*, which was then converted to forest with that probability. In practise, this was implemented as follows: for each *x*, draw a random number from a uniform distribution bounded at 0 and 1, convert to forest *x* if this number is less than $P_{for,x,t}$. After these forestation events were implemented, $P_{for,x,t}$ was calculated for every location *x* again, allowing for another round of land cover change. A record of the number of cells undergoing change as well as the pattern of forest *vs.* non-forest is kept at each time step, which allows us to calculate the annual rate of deforestation from 1950 through 2009.

Instead of relying on a single time transition (*e.g.* 2001-2002) to estimate past deforestation until 1950, we took advantage of having nine annual transitions (2001 through 2010) and, at each model iteration, we randomly selected an annual transition (2001-2002 or 2005-2006, for example), which was then used to estimate deforestation until 1950. Furthermore, because we ran the model 100 iterations, we get an ensemble of 100 projections based on different initial maps of forest/deforestation and parameterised models fitted for those transitions. This allows us to take into consideration that global land cover maps include uncertainty in their predictions of land cover change and direct map comparisons are not recommended (Friedl et al. 2002).

B.3) Annual and cumulative spatial validation of model predictions (2001-2009)

Model projections were validated against observed data by calculating three metrics on a pixel-by-pixel basis: perfect match (the pixel classification in the observed and predicted maps matches exactly), omission (deforestation is observed but not predicted on that pixel) and commission (deforestation is predicted but not observed in that pixel), for the period between 2001 and 2009, both annually and cumulatively. These metrics were chosen because, although it can be considered more rigid when compared to neighbourhood metrics such as the Kappa-family (Pontius & Millones 2011), it is more informative in showing how well the model predicted change. Furthermore, we only assess the observed and predicted change, not the resulting land cover maps avoiding validating pixels that do not change between the two maps, which would inevitably lead to high but unrealistic values of perfect match. What we aim to achieve here is to know how well the model predicted change. Given that different model iterations can start in different years (any year between 2002 and 2010), we did the validations based on model time (simulation time) rather than specific year. We want to understand if the model gains/losses predictive power through simulation years, on an annual and cumulative basis.

In addition, to assess if the model was able to predicted the spatial pattern of change accurately and, therefore, account for near and far misses (Pontius Jr et al. 2004b), we calculated a distance-based metric that allowed the quantification of the proportion of observed data that fell within a certain distance of the model predictions (1, 5, 10, 25, 50 km). In practise, we defined a buffer around each prediction and determined the number of observed change events found within that range, and then divided that number for the total change to obtain proportion.

B.4) Terrestrial carbon model

In order to estimate carbon losses from deforestation we combine the predictions from the spatial model with the plant carbon prediction maps resultant from Smith et al. (2012) model (Fig. B1). The authors developed a model based on data-constrained parameters of all component processes (*e.g.* plant carbon, leaf mortality, land burned, biomass allocation, fixation rate) within the global terrestrial carbon cycle. The great majority of these processes were found to have strong climate dependencies (Smith et al. 2012). Their terrestrial carbon model predicted global patterns of equilibrium plant carbon (with a measure of uncertainty) which were used in this study to determine the amount of carbon lost due to deforestation.

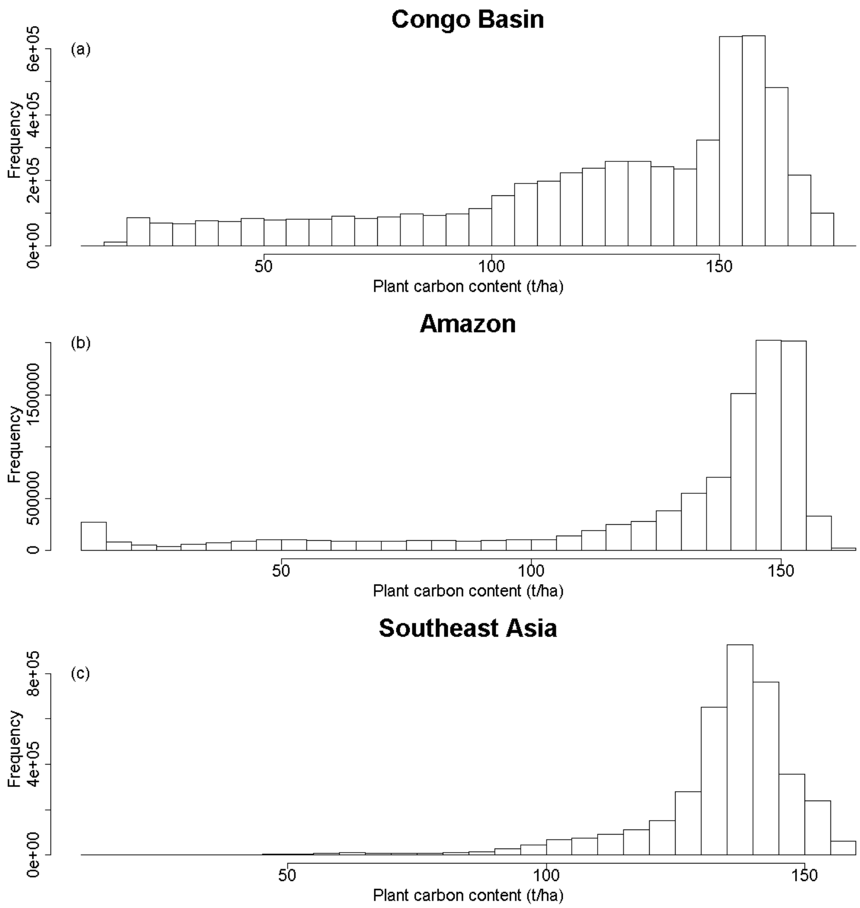


Figure B1 – Average plant carbon content (C t/ha) per pixel in (a) the Congo basin, (b) the Amazon and (c) Southeast Asia, using 100 maps derived from the initial median and confidence intervals carbon maps provided by Smith *et al.* (2012).

B.5) Carbon legacy from Southeast Asia

From SE Asia there would be 1.0 Pg C (± 0.16, 95% C.I.) to be released from the mainland, 0.55 Pg C (± 0.13, 95% C.I.) from the island of Sumatra, 0.44 Pg C (± 0.13, 95% C.I.) from the Borneo island, 0.29 Pg C (± 0.07, 95% C.I.) from the archipelago of Philippines, 0.06 Pg C (± 0.02, 95% C.I.) from the island of Sulawesi, and 0.06 Pg C (± 0.01, 95% C.I.) from the island of Papua. This would correspond to approximately 10 years of deforestation at the rate of 2000s in mainland (average 0.10 Pg C/yr), 8 years in Sumatra (average 0.07 Pg C/yr), 9 years in Borneo (average 0.05 Pg C/yr), 7 years in the Philippines (average 0.04 Pg C/yr), 6 years in Sulawesi (average 0.01 Pg C/yr), and 6 years in Papua (average 0.01 Pg C/yr).

Table B1 – Model parameters (mean ± 95% C.I.) for each annual transition between 2001 and 2010 for the Amazon region, which include Colombia, Peru, Venezuela, Suriname, French Guiana, Guyana, Bolivia, Ecuador and Brazil. Each parameter distribution is then used to sample a new value for each parameter to be used in at each model iteration.

Transition Parameter	2001 – 2002	2002 – 2003	2003 – 2004	2004 – 2005	2005 – 2006	2006 – 2007	2007 – 2008	2008 – 2009	2009 – 2010
Intercept	-4.22	-4.76	-4.40	-4.56	-4.30	-4.40	-4.45	-4.14	-4.29
	± 0.03	± 0.05	± 0.03	± 0.04	± 0.03	± 0.03	± 0.03	± 0.04	± 0.03
Neighbourhood	2.75	4.06	5.13	4.57	4.05	3.55	3.31	3.09	2.74
	± 0.07	± 0.06	± 0.04	± 0.04	± 0.04	± 0.04	± 0.04	± 0.03	± 0.04
Roads	0.31	0.11	0.39	0.22	0.39	0.62	0.98	1.07	1.29
	± 0.10	± 0.09	± 0.07	± 0.10	± 0.08	± 0.10	± 0.08	± 0.07	± 0.08
Rivers	-0.06	0.26	-0.15	-0.19	-0.70	-0.38	-0.64	-0.63	-1.11
	± 0.05	± 0.06	± 0.05	± 0.06	± 0.05	± 0.04	± 0.07	± 0.07	± 0.07
Altitude									
Dry season									
Pop. Density									
Protected	-0.02	0.14	0.20	0.06	0.22	0.38	0.48	0.35	0.30
Areas	± 0.05	± 0.05	± 0.05	± 0.05	± 0.04	± 0.04	± 0.05	± 0.04	± 0.05

Table B2 – Model parameters (mean ± 95% C.I.) for each annual transition between 2001 and 2010 for the Congo Basin region, which includes Cameroon, Equatorial Guinea, Congo Democratic Republic, Central Africa Republic, Gabon and Congo Republic. Each parameter distribution was then used to randomly sample a new value at each model iteration.

Transition Parameter	2001 – 2002	2002 – 2003	2003 – 2004	2004 – 2005	2005 – 2006	2006 – 2007	2007 – 2008	2008 – 2009	2009 – 2010
Intercept	-4.25 ± 0.07	-6.48 ± 0.30	-4.72 ± 0.12	-4.67 ± 0.50	-4.79 ± 0.08	-4.13 ± 0.88	-3.56 ± 0.19	-4.14 ± 0.05	-2.72 ± 0.13
Neighbourhood	4.58 ± 0.09	4.46 ± 0.11	4.75 ± 0.12	4.26 ± 0.14	4.73 ± 0.11	4.28 ± 0.17	3.77 ± 0.09	4.30 ± 0.12	2.75 ± 0.07
Roads	-0.13 ± 0.34	-1.80 ± 0.76	-3.33 ± 0.61	-0.75 ± 0.45	-1.34 ± 0.41	-0.79 ± 0.68	-0.52 ± 0.35	-2.29 ± 0.46	0.24 ± 0.26
Rivers	0.78 ± 0.07		-0.59 ± 0.15	0.17 ± 0.16	-0.55 ± 0.14	-0.53 ± 0.05	0.20 ± 0.17	-0.36 ± 0.05	-0.14 ± 0.05
Altitude	-0.001 ± 1×10^{-4}	-4×10^{-4} ± 3×10^{-5}	0.001 ± 4×10^{-5}	3×10^{-4} ± 4×10^{-5}	6×10^{-4} ± 3×10^{-5}	-2×10^{-5} ± 1×10^{-4}	-0.001 ± 8×10^{-5}	-2×10^{-5} ± 5×10^{-5}	-0.002 ± 8×10^{-5}
Dry season	0.0001 ± 3×10^{-5}	0.0001 ± 4×10^{-5}		5×10^{-5} ± 1×10^{-4}		3×10^{-5} ± 1×10^{-4}	-2.5×10^{-5} ± 5×10^{-5}	3×10^{-5} ± 6×10^{-5}	-2×10^{-5} ±4×10^{-5}
Pop. Density	0.001 ± 2×10^{-4}	0.42 ± 0.17	-0.11 ± 0.03	-0.04 ± 0.04	-0.0005 ± 2×10^{-4}	-0.08 ± 0.08	-0.0002 ± 0.002	2×10^{-5} ± 2×10^{-4}	0.16 ± 0.02
Protected Areas	0.05 ± 0.05	-0.26 ± 0.001	-0.25 ± 0.19	-0.27 ± 0.15	-0.44 ± 0.11	-0.11 ± 0.21	-0.19 ± 0.19	-0.31 ± 0.11	-0.23 ± 0.09

Table B3 – Model parameters (mean ± 95% C.I.) for each annual transition between 2001 and 2010 for the Borneo Island, which include Brunei and part of Malaysia and Indonesia. Each parameter distribution was then used to randomly sample a new value at each model iteration.

Transition Parameter	2001 – 2002	2002 – 2003	2003 – 2004	2004 – 2005	2005 – 2006	2006 – 2007	2007 – 2008	2008 – 2009	2009 – 2010
Intercept	-1.26	-3.62	-3.36	-3.63	-2.80	-3.21	-2.70	-2.89	-2.95
	± 0.08	± 0.12	± 0.10	± 0.09	± 0.09	± 0.09	± 0.07	± 0.08	± 0.10
Neighbourhood	0.96	2.74	3.09	3.48	2.95	3.51	2.85	3.16	2.69
	± 0.12	± 0.20	± 0.12	± 0.20	± 0.10	± 0.18	± 0.11	± 0.10	± 0.12
Roads	-0.52	-0.87	-0.09	0.21	0.47	0.53	1.37	0.73	0.14
	± 0.18	± 0.26	± 0.24	± 0.26	± 0.20	± 0.27	± 0.27	± 0.21	± 0.22
Rivers	-0.73	0.51	0.52	0.24	0.05	-0.13	-0.33	0.11	0.04
	± 0.07	± 0.22	± 0.14	± 0.14	± 0.14	± 0.05	± 0.10	± 0.09	± 0.13
Altitude	-0.0003					6×10^{-4}		7×10^{-4}	
	± 0.002					$\pm 2 \times 10^{-4}$		$\pm 1 \times 10^{-4}$	
Dry season	3×10^{-4}	4×10^{-4}	4×10^{-4}	2×10^{-4}	8×10^{-4}	-5×10^{-5}	8×10^{-6}	2×10^{-5}	-3×10^{-5}
	± 0.0001	$\pm 5 \times 10^{-5}$	$\pm 5 \times 10^{-5}$	$\pm 6 \times 10^{-5}$	$\pm 5 \times 10^{-5}$	$\pm 4 \times 10^{-5}$	$\pm 4 \times 10^{-6}$	$\pm 4 \times 10^{-6}$	$\pm 4 \times 10^{-6}$
Pop. Density	0.007	-0.26	0.01	2×10^{-4}	0.06	5×10^{-4}	0.006	$- 6 \times 10^{-4}$	-4×10^{-5}
	± 0.008	± 0.04	± 0.02	$\pm 2 \times 10^{-4}$	± 0.01	$\pm 2 \times 10^{-4}$	± 0.02	$\pm 3 \times 10^{-4}$	$\pm 2 \times 10^{-6}$
Protected	0.86	0.0098	-0.15	0.08	-0.05	0.03	0.02	-0.12	-0.09
Areas	± 0.08	± 0.13	± 0.13	± 0.09	± 0.13	± 0.05	± 0.11	± 0.10	± 0.05

Table B4 – Model parameters (mean ± 95% C.I.) for each annual transition between 2001 and 2010 for the mainland area of SE Asia, which includes peninsular Malaysia, Thailand, Cambodia, Vietnam and Laos. Each parameter distribution is then used to sample a new value at each model iteration.

Transition **Parameter**	**2001 – 2002**	**2002 – 2003**	**2003 – 2004**	**2004 – 2005**	**2005 – 2006**	**2006 – 2007**	**2007 – 2008**	**2008 – 2009**	**2009 – 2010**
Intercept	-4.94 ± 0.05	-5.57 ± 0.06	-4.74 ± 0.04	-5.39 ± 0.03	-5.02 ± 0.04	-5.69 ± 0.05	-6.17 ± 0.06	-5.73 ± 0.09	-5.11 ± 0.07
Neighbourhood	4.75 ± 0.05	4.83 ± 0.11	5.31 ± 0.07	5.44 ± 0.07	5.81 ± 0.08	5.92 ± 0.09	5.63 ± 0.1	5.93 ± 0.12	5.41 ± 0.10
Roads	-0.61 ± 0.001		1.43 ± 0.39	-1.18 ± 0.005			-0.11 ± 0.49	1.16 ± 0.98	0.31 ± 0.74
Rivers	0.61 ± 0.004	0.80 ± 0.06		1.18 ±0.006			0.21 ± 0.02	0.49 ± 0.23	0.17 ± 0.04
Altitude	-0.002 ± 2 $\times10^{-4}$	-0.0003 ± 9 $\times10^{-4}$	-0.001 ± 6 $\times10^{-5}$	-0.0001 ± 7 $\times10^{-4}$	-0.001 ± 8 $\times10^{-5}$	-0.0004 ± 6 $\times10^{-5}$	-0.0005 ± 5 $\times10^{-5}$	-0.001 ± 9 $\times10^{-5}$	-0.002 ± 4 $\times10^{-5}$
Dry season	3 $\times10^{-4}$ ± 4 $\times10^{-5}$		-0.001 ± 6 $\times10^{-5}$				0.0002 ± 4 $\times10^{-5}$	-5 $\times10^{-6}$ ± 0.0001	0.0003 ± 2 $\times10^{-5}$
Pop. Density	0.002 ± 1 $\times10^{-4}$	0.0003 ± 0.0002	0.001 ± 1 $\times10^{-4}$	0.0001 ± 0.0002	0.001 ± 0.0001	0.0004 ± 0.0001	0.0006 ± 0.0004	0.0003 ± 0.0001	0.002 ± 0.0001
Protected Areas	-0.17 ± 0.06	0.13 ± 0.001	0.25 ± 0.05	-0.18 ± 0.10	-0.04 ± 0.08	-0.49 ± 0.08	-0.15 ± 0.09	-0.30 ± 0.05	-0.03 ± 0.12

Table B5 – Model parameters (mean ± 95% C.I.) for each annual transition between 2001 and 2010 for the Papua Island, which comprises parts of Papua New Guinea and Indonesia. Each parameter distribution was then used to randomly sample a new value at each model iteration.

Transition Parameter	2001 – 2002	2002 – 2003	2003 – 2004	2004 – 2005	2005 – 2006	2006 – 2007	2007 – 2008	2008 – 2009	2009 – 2010
Intercept	-1.49 ± 0.13	-4.5 ± 0.24	-3.76 ± 0.18	-4.77 ± 0.18	-3.99 ± 0.25	-4.51 ± 0.30	-3.39 ± 0.17	-4.05 ± 0.17	-3.21 ± 0.20
Neighbourhood	1.64 ± 0.16	3.45 ± 0.62	3.74 ± 0.21	3.85 ± 0.18	3.75 ± 0.40	3.99 ± 0.35	2.81 ± 0.28	3.76 ± 0.16	2.45 ± 0.20
Roads		-1.07 ± 0.29	-0.60 ± 0.18	-0.35 ± 0.31	-0.09 ± 0.21	0.67 ± 0.18	0.009 ± 0.25	0.26 ± 0.20	0.49 ± 0.36
Rivers	-1.19 ± 0.06		0.18 ± 0.26	0.69 ± 0.06		-1.25 ± 0.04	0.54 ± 0.09	0.70 ± 0.37	0.27 ±0.03
Altitude	-0.001 ± 0.0001			0.001 ± 9×10^{-5}	0.0003 ± 6×10^{-4}	0.0004 ± 8×10^{-5}	2×10^{-5} ± 1×10^{-5}	5×10^{-5} ± 1×10^{-5}	4×10^{-4} ± 6×10^{-5}
Dry season						6×10^{-5} ± 4×10^{-5}			-0.001 ± 0.001
Pop. Density	-0.35 ± 0.03	-8×10^{-5} ± 4×10^{-5}						0.005 ± 0.01	
Protected Areas	0.49 ± 0.33	0.36 ± 0.25		-0.36 ± 0.37	-0.34 ± 0.55	-0.10 ± 0.41	-0.18 ± 0.35	-0.13 ± 0.34	-0.19 ± 0.37

Table B6 – Model parameters (mean ± 95% C.I.) for each annual transition between 2001 and 2010 for the Philippines archipelago. Each parameter distribution was then used to randomly sample a new value at each model iteration.

Transition Parameter	2001 – 2002	2002 – 2003	2003 – 2004	2004 – 2005	2005 – 2006	2006 – 2007	2007 – 2008	2008 – 2009	2009 – 2010
Intercept	-4.27 ± 0.07	-4.93 ± 0.08	-4.31 ± 0.07	-5.02 ± 0.07	-3.50 ± 0.06	-3.88 ± 0.07	-4.83 ± 0.09	-4.15 ± 0.08	-4.36 ± 0.07
Neighbourhood	2.94 ± 0.15	3.74 ± 0.16	4.24 ± 0.17	4.69 ± 0.12	4.86 ± 0.13	4.65 ± 0.14	4.76 ± 0.18	4.73 ± 0.12	4.37 ± 0.12
Roads		0.89 ± 1.20	-0.13 ± 0.003	3.49 ± 0.93	0.81 ± 0.78	0.78 ± 0.65			
Rivers	0.83 ± 0.14	0.73 ± 0.05	0.13 ± 0.001	0.75 ± 0.07	0.19 ± 0.10	0.13 ± 0.06			
Altitude		0.001 ± 9×10^{-5}	4×10^{-4} ± 6×10^{-5}	5×10^{-4} ± 1×10^{-4}	-0.001 ± 1×10^{-4}	1×10^{-4} ± 1×10^{-4}	5×10^{-4} ± 9×10^{-5}		-3×10^{-4} ± 8×10^{-6}
Dry season		6×10^{-5} ± 6×10^{-5}	-1×10^{-5} ± 2×10^{-6}	3×10^{-5} ± 8×10^{-6}		7×10^{-5} ± 6×10^{-5}	-3×10^{-5} ± 7×10^{-6}		
Pop. Density			0.007 ± 0.06		-0.23 ± 0.02	-0.19 ± 0.01	-5×10^{-4} ± 3×10^{-4}		4×10^{-5} ± 2×10^{-5}
Protected Areas	0.06 ± 0.17	-0.06 ± 0.14	0.015 ± 0.002	-0.13 ± 0.16	0.05 ± 0.15	-0.005 ± 0.13			-0.29 ± 0.09

Table B7 – Model parameters (mean ± 95% C.I.) for each annual transition between 2001 and 2010 for the Sulawesi Island. Each parameter distribution was then used to randomly sample a new value at each model iteration.

Transition / Parameter	2001 – 2002	2002 – 2003	2003 – 2004	2004 – 2005	2005 – 2006	2006 – 2007	2007 – 2008	2008 – 2009	2009 – 2010
Intercept	-3.47 ± 0.09	-4.48 ± 0.20	-3.44 ± 0.10	-4.30 ± 0.20	-3.80 ± 0.10	-4.27 ± 0.17	-4.13 ± 0.12	-4.31 ± 0.15	-4.72 ± 0.22
Neighbourhood	2.69 ± 0.25	3.33 ± 0.37	4.08 ± 0.21	3.66 ± 0.23	4.17 ± 0.19	3.86 ± 0.26	3.18 ± 0.27	4.21 ± 0.24	3.93 ± 0.40
Roads	2.57 ± 1.91	3.22 ± 1.63	3.02 ± 1.72	4.46 ± 0.49	1.42 ± 1.89	4.33 ± 0.64	2.18 ± 1.99	4.21 ± 0.77	3.61 ± 1.17
Rivers	-0.30 ± 0.07			0.10 ± 0.18					
Altitude	$2\ x10^{-4}$ ± $7x10^{-5}$	$7\ x10^{-6}$ ± $1\ x10^{-4}$	$2\ x10^{-4}$ ± $2\ x10^{-5}$	$3\ x10^{-4}$ ± $2\ x10^{-4}$	$3\ x10^{-4}$ ± $2\ x10^{-4}$	$7\ x10^{-5}$ ± $2\ x10^{-5}$	$5\ x10^{-4}$ ± $1\ x10^{-4}$	$2\ x10^{-4}$ ± $1\ x10^{-4}$	$-5\ x10^{-4}$ ± $3\ x10^{-4}$
Dry season	$-2\ x10^{-4}$ ± $1x10^{-4}$		$-2\ x10^{-4}$ ± $1\ x10^{-4}$		$-1.5\ x10^{-5}$ ± $1\ x10^{-5}$				
Pop. Density	$-\ 4\ x10^{-4}$ ± $1x10^{-4}$		-0.07 ± 0.02	-0.26 ± 0.06	$-5\ x10^{-5}$ ± $2\ x10^{-5}$	0.004 ± 0.008	$-5\ x10^{-4}$ ± $5\ x10^{-5}$	$-1\ x10^{-4}$ ± $1\ x10^{-5}$	$5\ x10^{-4}$ ± $4\ x10^{-4}$
Protected Areas		0.02 ± 0.28		0.07 ± 0.32	-0.32 ± 0.14		-0.08 ± 0.10		

Table B8 – Model parameters (mean ± 95% C.I.) for each annual transition between 2001 and 2010 for the Sumatra Island. Each parameter distribution was used to randomly sample a new value at each model iteration.

Transition Parameter	2001 – 2002	2002 – 2003	2003 – 2004	2004 – 2005	2005 – 2006	2006 – 2007	2007 – 2008	2008 – 2009	2009 – 2010
Intercept	-1.50	-3.68	-3.23	-3.84	-3.19	-3.63	-2.96	-3.47	-2.97
	± 0.05	± 0.09	± 0.08	± 0.08	± 0.07	± 0.05	± 0.05	± 0.08	± 0.05
Neighbourhood	1.72	3.05	3.19	4.03	3.69	4.08	3.56	3.88	3.68
	± 0.06	± 0.13	± 0.12	± 0.11	± 0.08	± 0.09	± 0.10	± 0.09	± 0.09
Roads	1.75	0.39	-0.31	0.04	0.83		0.40	0.10	0.40
	± 0.18	± 0.34	± 0.41	±0.32	± 0.24		± 0.25	± 0.29	± 0.43
Rivers	-1.12		0.27	0.25		0.01	0.06	0.26	-0.49
	± 0.05		± 0.06	± 0.04		± 0.05	± 0.04	± 0.03	± 0.05
Altitude	-0.001	-0.0001	2×10^{-4}	0.0003	0.0002	-0.0003	-0.0004	0.0001	-8×10^{-4}
	$\pm 7 \times 10^{-4}$	$\pm 9 \times 10^{-5}$	$\pm 5 \times 10^{-5}$	$\pm 6 \times 10^{-4}$	$\pm 8 \times 10^{-4}$	$\pm 7 \times 10^{-4}$	$\pm 4 \times 10^{-5}$	$\pm 6 \times 10^{-5}$	$\pm 8 \times 10^{-5}$
Dry season	2×10^{-6}		2×10^{-5}	4×10^{-4}			-0.0005	-2×10^{-4}	5×10^{-5}
	$\pm 3 \times 10^{-6}$		$\pm 6 \times 10^{-6}$	$\pm 3 \times 10^{-5}$			± 0.0001	$\pm 7 \times 10^{-5}$	$\pm 1 \times 10^{-5}$
Pop. Density	0.001	0.002	-0.20	-0.0002	-0.13	0.005	0.0002	1.5×10^{-4}	8×10^{-4}
	± 0.0001	± 0.004	± 0.03	± 0.0004	± 0.03	± 0.008	± 0.0002	$\pm 2 \times 10^{-5}$	$\pm 2 \times 10^{-4}$
Protected Areas	-0.11	0.002	0.12	0.06	0.16	0.11	-0.03	-0.01	0.16
	± 0.0001	± 0.001	± 0.09	± 0.04	± 0.09	± 0.14	± 0.09	± 0.008	± 0.04

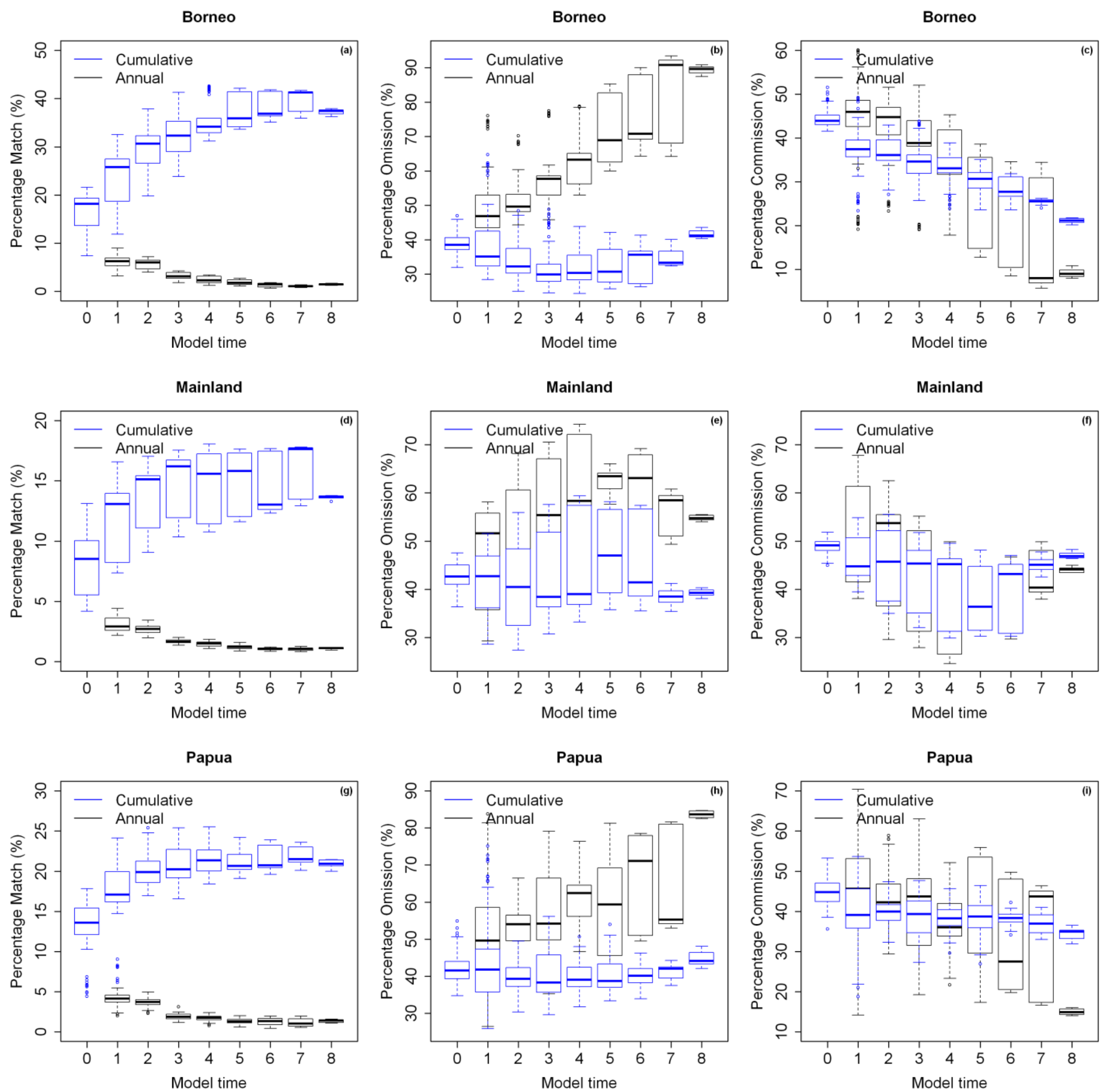


Figure B2 – Pixel by pixel validation results for the first 9 years of model predictions for Borneo, Mainland and Papua models: (a), (d), (g) show perfect match; (b), (e), (h) show omission and (c), (f), (i) show commission).

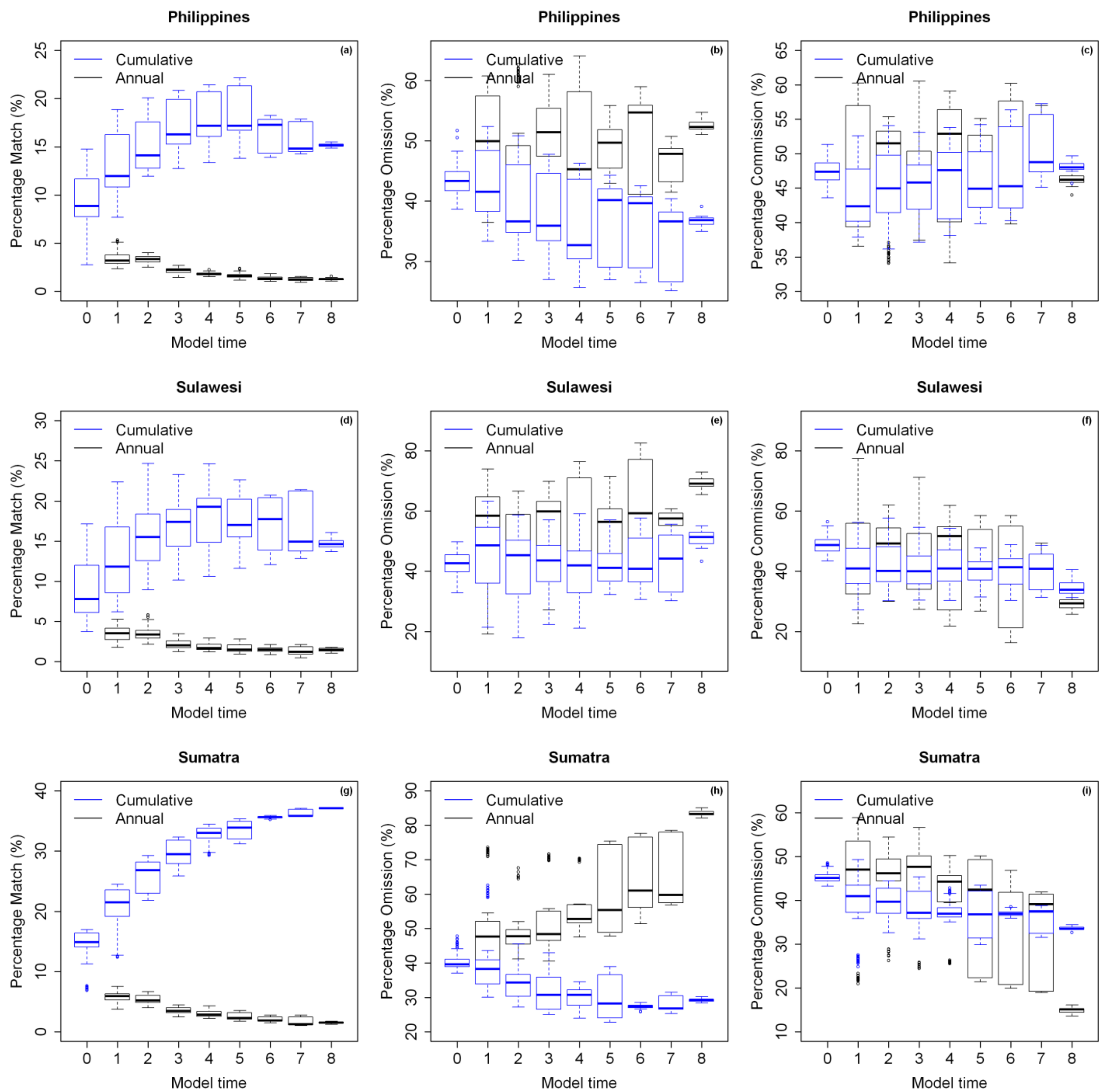


Figure B3 – Pixel by pixel validation results for the first 9 years of model predictions for Philippines, Sulawesi and Sumatra models: (a), (d), (g) show perfect match; (b), (e), (h) show omission and (c), (f), (i) show commission).

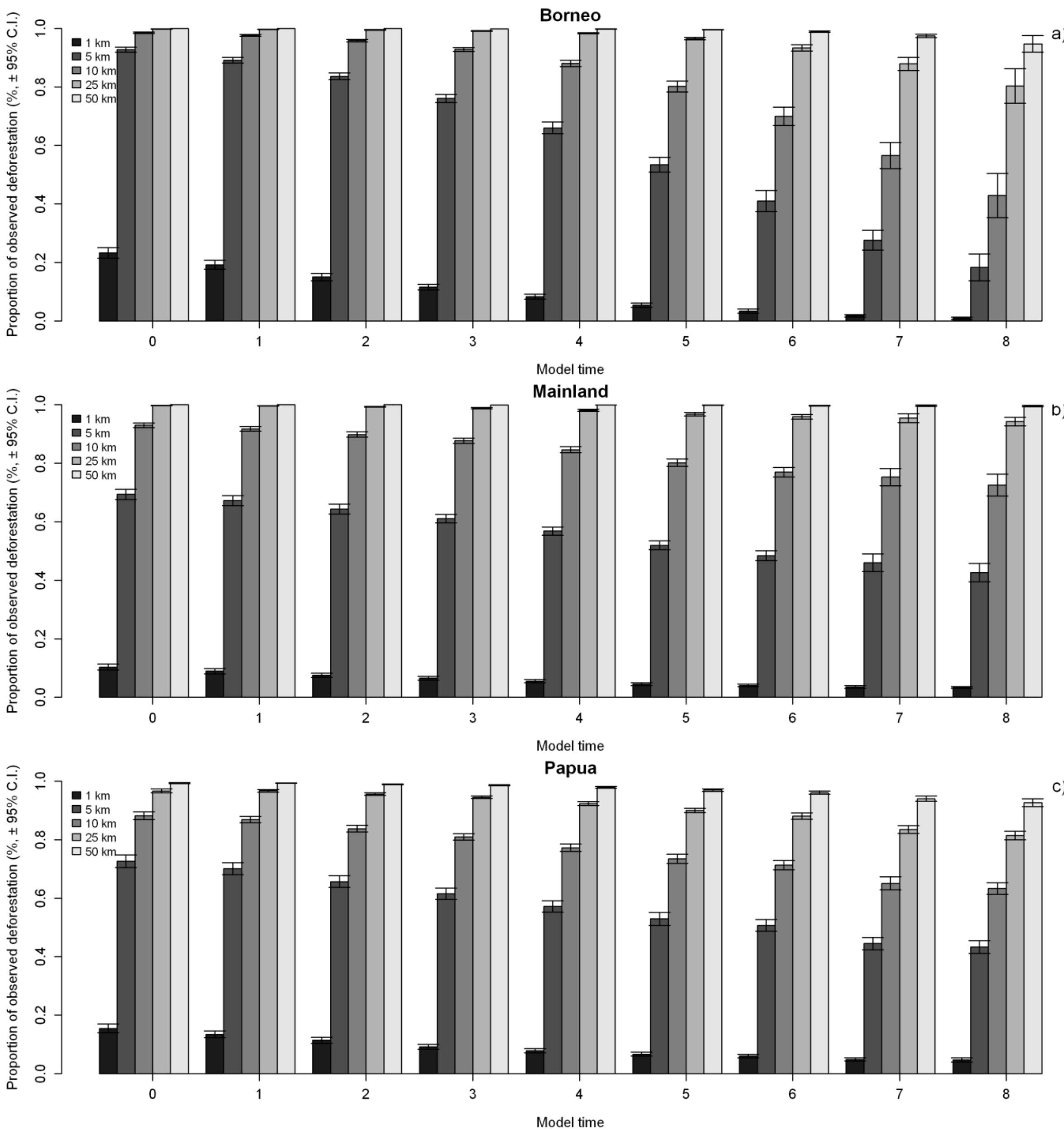


Figure B4 – Distance-based metric of validation: proportion of observed deforestation within certain distances of model predictions in (a) Borneo, (b) Mainland and (c) Papua through the first 9 years of model predictions.

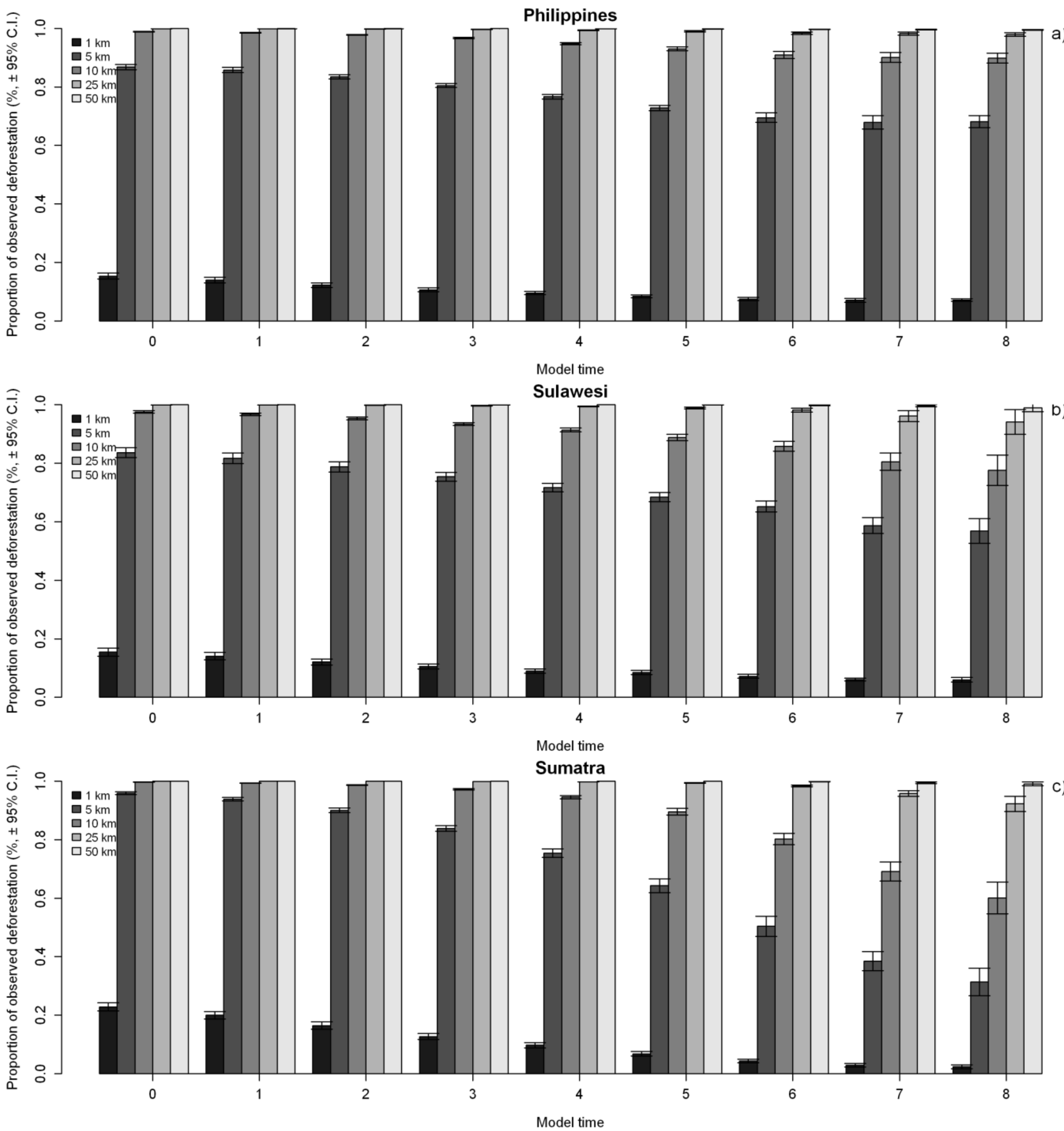


Figure B5 – Distance-based metric of validation: proportion of observed deforestation within certain distances of model predictions in (a) Philippines, (b) Sulawesi and (c) Sumatra through the first 9 years of model predictions.