Road Development in the Brazilian Amazon and its Ecological Implications

Sadia Evelyn Ahmed

June 2013

A thesis submitted for the degree of Doctor of Philosophy, Diploma of Imperial College (DIC) from the Division of Ecology and Evolution, Department of Life Sciences, Imperial College London

Declaration of originality

I hereby declare that all work submitted in this thesis is my own original research. Where information has been derived from other sources, these have been cited and referenced appropriately. All collaborative work has been properly acknowledged.

Sadia Evelyn Ahmed

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution-Non Commercial-No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or distribution, researchers must make clear to others the licence terms of this work.

Abstract

Roads are a distinctive feature in any landscape, with many countries giving 1-2% of their land surface over to roads and roadsides (Forman 1998). However, the ecological effects of roads spread beyond the physical footprint of the network and may impact 15-20% of the land or more (Forman & Alexander 1998).

The Brazilian Amazon contains approximately one third of the world's remaining rainforest, covering an area of 4.1 million km^2. The region is highly biodiverse with 10-20 percent of the planet's known species, it is also one of the three most bioculturally diverse areas in the world (Loh & Harmon 2005), and it provides many valuable ecosystem services. However, the Brazilian Amazon is rapidly undergoing extensive development with widespread land-use conversion.

Road development is often perceived as the initial stage of development, opening access to remote areas for colonisation, agriculture development, resource extraction, and linked with these; deforestation (Chomitz & Gray 1996, Laurance *et al.* 2001, Perz *et al.* 2007, Laurance *et al.* 2009, Caldas *et al.* 2010). As such roads are a key spatial determinant of land use conversion in the Amazon region, dictating the spatial pattern of deforestation and biodiversity loss (Fearnside 2005, Kirby *et al.* 2006, Perz *et al.* 2008).

Given that roads are a key spatial determinant of land use conversion and that they have extensive impacts on rates and patterns of habitat loss, it is important that we know how much, how fast and where road networks are developing in this globally important ecosystem. In this thesis, I aim to construct models of road network development to help better understand and predict the impacts of economic development in the Brazilian Amazon.

Acknowledgments

> *"Here no one who has any feeling of the magnificent and the sublime can be disappointed; the sombre shade, scarce illumined by a single direct ray even of the tropical sun, the enormous size and height of the trees... the strange buttresses around the base of some, the spiny or furrowed stems of others, the curious and even extraordinary creepers and climbers which wind around them, hanging in long festoons from branch to branch... the wonderful variety of the foliage, the strange fruits and seeds... the rarest birds, the most lovely insects, and the most interesting mammals and reptiles are to be found. Here lurk the jaguar and the boa-constrictor, and here amid the densest shade the bell-bird tolls his peal."*
>
> **Alfred Russel Wallace on the Amazon, 1849**

I would like to thank my supervisors Robert Ewers and Drew Purves for giving me the opportunity to pursue this PhD. I would especially like to thank Rob for all his valuable input to my work, for providing advice and assistance when I needed it, whilst allowing me to work in my own way. I would also like to thank Matthew Smith for all his time, support and encouragement. David Orme for his invaluable help, Simon Leather for always being there to talk to, Taru Khana for innumerable I.T. fixes, & Diana Anderson for dealing with admin and being brilliant. I would also like to take this opportunity to thank ALL the Lecturers and Professors at Imperial, who have taught and inspired me, from the first day of my first degree; thank you for shaping the academic I hope to become.

This research would not have been possible without the sponsorship of Microsoft Research Cambridge and the Grantham Institute for Climate Change.

Thank you to my parents for their support, especially to Mummy. Thank you to my siblings, Shahrukh, Sana & Sarrinah. Thanks to Skuncle for making me sit still and watch Attenborough. I suppose biology was a foregone conclusion! And.....

Thank you to my husband Matt, without whom I would not have survived. You deserve a whole page to yourself.

Thank you for all your support, love and friendship. Thank you for accepting me for who I am. Thank you for your understanding and patience. Thank you for always being here when I need you.

Yours always & forever

X

Contents

List of figures

which development is moving is much more pronounced when four regions are considered in contrast to the Amazon as a whole (Figure 5.1).

List of tables

List of appendices

Chapter 1: **Introduction**

1.1. Introduction

Roads are an important and necessary part of everyday life for most people, forming the basis of the overland transportation network (along with railways) in nearly all countries. Road development influences a wide range of phenomena, from human society, business and economies, to the natural environment (Forman *et al.* 2003). In regional development, roads are often perceived as the initial stage of development, especially in tropical areas where they open access to remote areas for colonisation, agricultural development, and resource extraction (Laurance *et al.* 2001, Arima *et al.* 2005, Perz *et al.* 2007, Caldas *et al.* 2010). Roads further facilitate development by providing market access for rural producers, integrating economic sectors and reducing the cost of spatial mobility (Perz *et al.* 2007).

Global road networks have been expanding at a rapid rate since the 1900's (Forman *et al.* 2003), making roads a distinctive feature in any landscape, with many countries giving 1-2% of their land surface over to roads and roadsides (Forman 1998). In many emerging economies, road building is vital for stimulating and maintaining economic growth (Andersen & Reis 1997). In Brazil for example, infrastructure initiatives have been used since the 1970s to this end (Carvalho *et al.* 2002, Alves 2002, Kirby *et al.* 2006, Ahmed *et al.* 2013), indeed such initiatives are still in use today in the region, for example the Initiative for the Integration of the Regional Infrastructure of South America (IIRSA) project (Killeen 2005). Today the highest rates of road expansion can be seen in the developing tropics and in emerging economies, where roads are given high priority by governments to encourage growth and reduce poverty through increasing spatial connectivity, aiding travel, helping establish land claims and facilitating the extraction of resources (Munnell 1992, Calderon & Serven 2004, Straub 2008, Perz *et al.* 2012). Indeed roads are expanding at rapid rates across the tropics, for example, on average 17,000 km of roads are added to the Brazilian Amazon each year (Ahmed et al. 2013).

Despite the irrefutable socio-economic benefits that roads bring to humans, they often result in negative impacts on the environment (Forman & Alexander 1998, Spellerberg 1998, Fahrig & Rytwinski 2009, Laurance *et al.* 2009, Perz *et al.* 2012). The ecological effects of roads spread far beyond the physical footprint of the network and may impact 15-20% of the surrounding land (Forman & Alexander 1998). The ecological effects of roads are diverse, ranging from road mortality events, loss of habitat, the formation of barriers to animal dispersal and gene flow, to, altering habitat structure, creating edges, introducing pollutants, changing hydrological processes and increasing susceptibility to alien invasion (Forman & Alexander 1998, Keller & Largiader 2003, Laurance *et al.* 2004, Shyama Prasad Rao & Saptha Girish 2007, Jaeger *et al.* 2005). These effects vary across biomes, habitats and scales. Many road impacts eventually cause changes to biodiversity richness and species composition (Wilkie *et al.* 2000, Forman *et al.* 2003, Spooner & Smallbone 2009).

One of the most striking road effects is the impact roads have on deforestation in tropical regions. In the context of tropical deforestation, roads cause a relatively small amount of direct habitat loss, but exert a huge indirect influence on the spatial patterns of deforestation by allowing easier access to new frontiers (Fearnside 2008, Geist & Lambin 2002, Perz *et al.* 2007, Perz *et al.* 2008). Roads also encourage extractive industries and further deforestation by settlers, thereby indirectly influencing deforestation rates.

The close links between roads and deforestation means that roads are often a key spatial determinant of land use conversion; strongly influencing the rates and patterns of habitat loss. As such, infrastructure, including road and rail networks, is often incorporated into land use change models. These models project future land conversions with a view to quantify future changes in carbon flux, climate change and biodiversity. However, the spatio-temporal patterns of road network development are poorly understood and seldom quantified. It has been found that approximately two thirds of papers predicting land use change in the Amazon region use roads as a predictor of future land use (Rosa *et al.*2014). Yet, the majority of these land use change models treat road development as a static phenomena (given the rate at which roads change, this is simply not realistic). Thus models that can characterise and predict road development play a vital role in future land use modelling. As such, there is burgeoning interest in predicting road development especially the case of developing nations, which are high in natural resources, where road development is rapid and often not centrally managed. Unfortunately, characterisation of large scale spatiotemporal patterns in road network development has been greatly overlooked to date.

In the Brazilian Amazon the majority of roads built are unofficial and there is a distinct lack of spatial information on the location and extent of these roads (Brandão & Souza 2006). This presents a problem for policy makers and conservationists who need spatial information on current and future roads in order to assess potential impacts and make informed decisions. Given there are complex dynamics and interactions of road development with economics, policy, technology, demographic and cultural factors, which vary between regions, it is unsurprising that few models of road development exist to help the situation.

The Amazon rainforest is highly diverse, productive and offers vast array of ecosystem goods and services. It is subject to many pressures including, extractive industry, resource exploitation, poor governance, a changing climate and infrastructure development. Given the importance of this ecosystem, the fact that it has a rapidly growing road network (that is largely unplanned), and the extensive negative ecological effects a road network can cause, it is imperative that we understand and are able to predict the spatio-temporal dynamics of road network growth in this globally important system. To date four different road models have been used to predict the growth of the Brazilian road network (Arima *et al.* 2008, Soares-filho *et al.* 2004, Jiang 2007, Walker *et al.* 2013) all of which utilise least-cost path algorithms to determine the path of developing roads, however only two of these have been validated (Arima *et al.* 2008, Walker *et al.* 2013). It is hoped that the results of this thesis can be used to further improve our understanding and ability to predict road development and its ecological implications in this valuable ecosystem. The thesis focuses on road development in the Brazilian Amazon.

1.2. Thesis outline

The main objectives of this thesis were (1) to investigate the ecological effects of roads, (2) to generate amazon-wide road models that could potentially be used to estimate the future of road development and their impacts in this globally important biome, and (3) to determine if these models would stand up to critical validations. To this end I have conducted a detailed literature review of ecological road effects (Chapter 1), assessed the potential of using road maps to estimate biodiversity (Chapter 2) and, developed two distinct and validated modelling methodologies. In this thesis I present a data constrained statistical model of road development (Chapter 5) and a process based model (Chapter 6) building on the concepts presented by Arima et al. (2008), in addition to two other chapters that provide integral preliminary analyses for both models. In order to model future road development, past development needs to be understood. Therefore, in addition to the qualitative description of road development in the Amazon presented in Chapter 1, Chapter 4 quantifies past patterns of development. The process based model assumes initial roads into forests are logging roads, and given loggers would seek to maximise profits, it assumes that roads will tend towards high revenue timber stands. Chapter 3 generates the revenue map that the process based model relies on.

Chapters 3, 4 & 5 have been submitted to peer-reviewed journals with contributions from co-authors (as outlined below); consequently first person plural is used. Each chapter, except the introduction and discussion, is written as a manuscript for submission to a journal. A consistent format has been adopted for the thesis irrespective of individual journal formatting.

Chapter 2: The ecological effects of roads

The first chapter introduces roads, road building and the ecological effects of roads, before going on to introduce the study region, the Brazilian Amazon, and road development as is specific to this region. This chapter aims to provide a thorough review and grounding of the thesis subject.

Chapter 3: Can roadless volume predict patterns of biodiversity? A test using birds in the central Amazon

Sadia E. Ahmed, Alexander C. Lees, Nárgila G. Moura, Toby A. Gardner, Jos Barlow, Joice Ferreira & Robert M. Ewers

In preparation for: Biological Conservation

A key implication of road development is the alteration of local biodiversity. In this chapter I aim to determine if (1) there is a relationship between roadless volume and biodiversity, and (2) what would predictions of species richness look like extrapolated from this relationship. Using bird species richness data I investigate if a metric of road network coverage, roadless volume, can be used to estimate species richness and if road networks influence bird community composition. Estimates of biodiversity loss as the road network has grown between 2000 and 2008 are also made for the study region.

The bird richness data from Santarém and Belterra in Pará state, eastern Brazilian Amazon, was collected by Alexander C. Lees, Nárgila G. Moura and a host of field assistants, and kindly given to me by the RAS (Rede Amazônia Sustentável) project steering committee (Toby A. Gardner, Jos Barlow & Joice Ferreira). Thanks are due to my co-authors for helpful comments on previous drafts of this manuscript.

Chapter 4: Spatial pattern of standing timber value across the Brazilian Amazon

Sadia E. Ahmed & Robert M. Ewers

Accepted: Plos ONE, March 2012

Given the importance of logging operations in the building of roads in the Amazon region, and that logger's build roads to access valuable timber. Here I aim to determine if there is a pattern to the distribution of timber value across the Amazon forest and if this pattern can be related to ecological processes. This chapter presents a map of where the most valuable timber stands in the Amazon are and further presents an ecological explanation for the observed distribution of value. Results from this chapter are later utilised in Chapter 6.

With thanks to Robert and two anonymous reviewers, whose comments greatly improved this manuscript.

Chapter 5: Temporal patterns of road network development in the Brazilian Amazon

Sadia E. Ahmed, Carlos M. Souza Jr, Júlia Riberio & Robert M. Ewers

Accepted: Regional Environmental Change, December 2012

In this chapter I aim to determine (1) what temporal dynamics of road density look like and (2) over what time scales the phases of road development occur. In addition to this, (3) if the observed pattern can be related to anthropogenic and economic phenomena. Past spatio-temporal patterns of road development in the Amazon are estimated and described using a space-for-time approach. This approach assumes road development is moving forward through the Amazonian arc-of-deforestation such that areas in front of the arc are likely to develop in a similar fashion as those behind the arc.

Amazon wide maps for two time points were kindly provided by Carlos M. Souza Jr. and Júlia Riberio. Data from this chapter are used to calibrate the model presented in Chapter 6. The comments from Robert and several anonymous reviewers' on this manuscript are wholly appreciated.

Chapter 6: Large scale spatiotemporal patterns of road development in the Amazon rainforest

Sadia E. Ahmed, Robert M. Ewers & Matthew J. Smith

Accepted*: Environmental conservation, September 2013 (*shortened version)

In light of the similarity between the growth dynamics seen in Chapter 4 and known population growth dynamics, a trans-disciplinary approach was taken in this chapter. By applying simple population models to the field of landscape ecology to model road dynamics. In this chapter I aim to determine if road development occurs in a directional manner and if model projections of road density development can be related to what is known of other development processes in the region, such as deforestation. I also seek to establish if the processes governing road dynamics are intrinsic only, or, if neighbourhood effects also play a role. 16 models of spatio-temporal road density development are assessed. Spatial correlation in density changes within municipios, and local vs. neighbourhood density influences are also examined. Further, the influence of barriers to development is incorporated in projections of the future development of road density across the Amazon for a 60 year period.

Gratitude is due here to Matthew, who coded the models in C#, was instrumental in the evaluation process and who read previous drafts of this manuscript; providing many useful comments.

Chapter 7: A process based model of road development in the Amazon

Sadia E. Ahmed, David C. Orme & Robert M. Ewers

Work in progress

In this chapter I aim to determine if a process based model utilising cost-revenue mechanisms can accurately predict road networks. A process-based model of road growth that predicts locations of individual roads, as opposed to overall road density (as in the preceding chapter) is presented and validated. This is necessary because due to the wide spread use of roadmaps in land use change modelling the models in Chapter 5 are not applicable, therefore in this chapter I sought to develop an actual ‘roadmap’ model. Here the assumption that logger’s seek valuable timber is used to select road destinations based on the results of Chapter 3. This chapter has a solely methodological focus and seeks to present the current state of the process based model.

I am extremely grateful to David for his help with Python and to Robert for reading previous drafts of this manuscript.

Chapter 8: Discussion

As each chapter has an individual discussion the final discussion does not deal with all individual results, rather it offers an overview and discussion of the thesis as a whole.

Appendix A: Model code for Chapter 6

Appendix B: Additional figure for Chapter 6

Appendix C: Model code for Chapter 7

Appendix D: The transparency, reliability and utility of land-use and land-cover change models: an Amazonian case study

Isabel M. D. Rosa*, Sadia E. Ahmed* & Robert M. Ewers (*joint first)

Accepted: Global Environmental Change, December 2013

This manuscript is a quantitative review of 35 modelling methodologies, considering model spatio-temporal scales, inputs, calibration and validation methods. In addition, a quantitative assessment of model performance for LULC predictions in the Brazilian Amazon was carried out for some of the models. Shortfalls in the discipline and three key points that need addressing to improve the transparency, reliability and utility of LULC change model are highlighted. While this work does not exactly fit this thesis remit, it was carried out during my PhD and is referenced several times. Thus, it has been included as an appendix for reference.

Deepest thanks to my co-author and friend, Isabel, are due here.

Chapter 2: **The ecological effects of roads**

2.1. A brief history of roads

Roads are an important and necessary part of everyday life for most people. They form the basis of the overland transportation network (with the railways) in nearly all countries. Roads have aided travel and trade, connected people and places, helped the rise of empires and economies, and shaped human landscape and history.

The first roads were footpaths and dirt trails which facilitated the movement of people, their goods and animals (Belloc 1923, Gregory 1931, Hindley 1971, Forman *et al.* 2003). Today roads serve the same function, but have developed into larger, more extensive networks of manufactured, hard surfaces. Originally, given that horse, oxen or walking were the main modes of transport, open land would not need any roads; roads were only needed along tougher terrain such as mountains, swamps and forests (Gregory 1931, Forbes 1964). The development of these early 'roads' was relatively simple; the safest path, the path of least resistance or the path currently most commonly used would be cleared of obstacles e.g. boulders and trees, giving rise to 'paths'(Forbes 1964, Hindley 1971). As these were used more they would naturally widen from traffic flow in the form of foot traffic, wagons, sleds, and beasts of burden, giving rise to 'tracks', 'ridge-ways' (which differ from other tracks because they were formed along the tops of hills, where soil tends to be hard and dry) and 'causeways' (which differ from other simple tracks because they are elevated on a sand bank, usually found in wet or boggy areas) (Forbes 1964). These basic paths and tracks can be collectively called 'unimproved roads'. As transportation developed so did roads, with developments in ground levelling and surface sealing giving rise to 'manufactured roads', also called 'improved roads' or 'paved roads' that are passable in most weather and suitable for more advanced transportation.

The earliest manufactured roads appeared approximately 5000 years ago in ancient Crete, Mesopotamia and the Himalayas (Gregory 1931, Forbes 1964, Hindley 1971). Despite several notable developments in road building taking place since, such as Roman roads (Gregory 1931), manufactured road engineering was only greatly improved during the 18th and 19th centuries. However, it was the railways, telegraphs and telephones that first overcame long distances (Forman *et al.* 2003). Indeed it was not until the 20th century, with the advent of the motor vehicle, that significant investment and development was made in road networks outside of cities (Forman *et al.* 2003). Since the 20th century, road networks around the globe have been expanding and have become a necessity of society and the economy. Today unimproved roads are still in use in areas where climatic and seasonal conditions reduce the validity of a manufactured road or in areas unable to afford the construction and maintenance of manufactured roads. Most countries however have a network of manufactured roads (Figure 2.1, World Bank 2011).

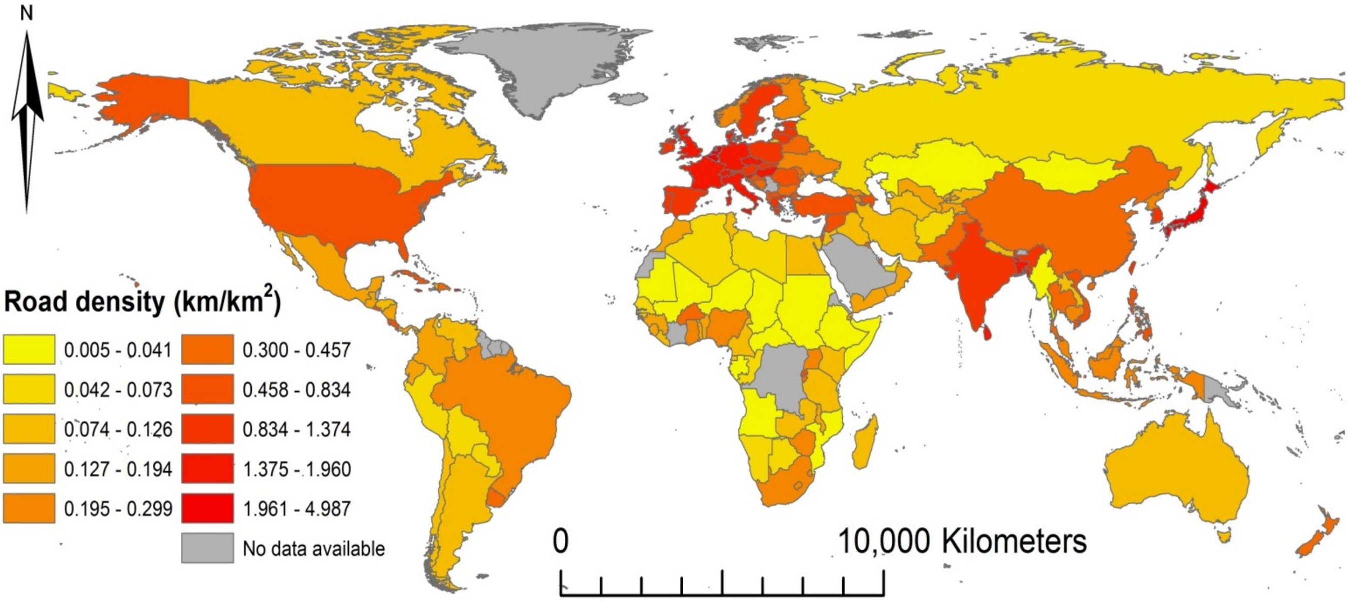


Figure 2.1. The global density distribution of manufactured roads.

2.2. Types of roads

Roads are constructed for many and varied purposes, giving them a wide range of political, topological and morphological differences. Official roads are built either by the government or with government permission, whereas unofficial roads are built with no planning permission obtained from the state by non-state actors, such as miners (Brandão & Souza 2006, Perz *et al.* 2007). Legal roads are any roads that are built within the limits of the law; they may or may not be government approved but they do not infringe any laws. By contrast, illegal roads break the law in some way, for example roads built in nature reserves where road building is forbidden. The political status of the road builder determines another important difference; federal roads are built with government funding and are under government control and maintenance, whereas private roads are built and maintained by private investors, be they businesses or private citizens.

Further divisions of roads exist based on morphological differences. The simplest way to divide roads based on morphology is paving, roads may be paved or unpaved. Paved roads have their surface covered with concrete, tarmac or another sealant, that makes the road more durable and gives an all weather surface. Unpaved roads lack a top layer sealant and tend to be temporary, often built for short term extraction projects, or the road is unfinished and there is the intention to pave over the road at a later date, such as the BR-163 highway in the Brazilian Amazon (Soares-Filho *et al.* 2004). Paved roads are by nature considered permanent because once paved they are long lasting, unlike unpaved roads that are essentially cleared, compacted dirt roads.

Road order is strongly influenced by the reason for building and refers to the road type in relation to its importance in the road network. There are three main orders of roads; primary (infrastructure), secondary and tertiary roads, with the later two being considered feeder or access roads. Primary roads are built to increase connectivity across a country or region, and include major highways. They form connections at the largest scales and are generally built with the intention to increase connectivity between urban centres. Although all roads form part of an infrastructure, the term infrastructure road is sometimes used to refer to the primary roads of a network. Secondary roads are the smaller roads that connect places within regions, and increase access to rural areas. Tertiary roads are smaller again and branch off secondary roads, further increasing the overall connectivity of a road network at a relatively fine spatial scale. Roads built specifically to access resources such as ores and timber, may give rise to secondary and tertiary roads that increase the overall road network, but can be called mining and logging roads respectively (based on building purpose instead of on order).

The emergence of a road network is an incremental process. Network development usually starts in urban or residential areas (be it dirt tracks connecting village huts or paved roads connecting town houses). The next step is to connect main urban areas and ports using primary roads (also called 'trunk' or 'regional' roads). This basic network is then expanded with secondary roads (also called 'cross roads') that connect regional roads and tertiary roads (also called 'access' roads) that provide access to resources (e.g. farms and timber) and allow transport from the resource source to the main road network. This general process and pattern of road building has been well documented (Taaffe *et al.* 1963).

Further to road types there are different types of road network pattern (Figure 2.2). These network patterns are morphologically comparable to the basic classes of river basin drainage networks. The six key drainage patterns are (1) dendritic, (2) rectangular, (3) trellis, (4) radial, (5) parallel, (6) annular, with each of these patterns determined by various geological factors (Zernitz 1932). Similarly, there are four broad categories of road networks and the different network patterns are often associated with different types of development. Radial patterns occur in primary roads that lead to and from a focal centre, such as a city (Figure 2.2a). Rectangular or gridded patterns (Figure 2.2b) are most commonly associated with either a planned settlement (where roads are laid in strait lines around 'blocks' of buildings) or with agricultural land (where roads boarder fields), explaining why this pattern has also been called a 'large property pattern' (Arima *et al.* 2008). Dendritic or organic road network patterns (Figure 2.2c) look similar to tree branches and are associated with 'unplanned' road development and extractive industries such as logging (Arima *et al.* 2008). Finally, fishbone road patterns (Figure 2.2d) correspond most closely to trellis drainage patterns and are most often associated with centrally planned human colonisation (Arima *et al.* 2005, Perz *et al.* 2007). Fishbone patterns are characterised by straight, relatively evenly placed, secondary roads leading off a central main road.

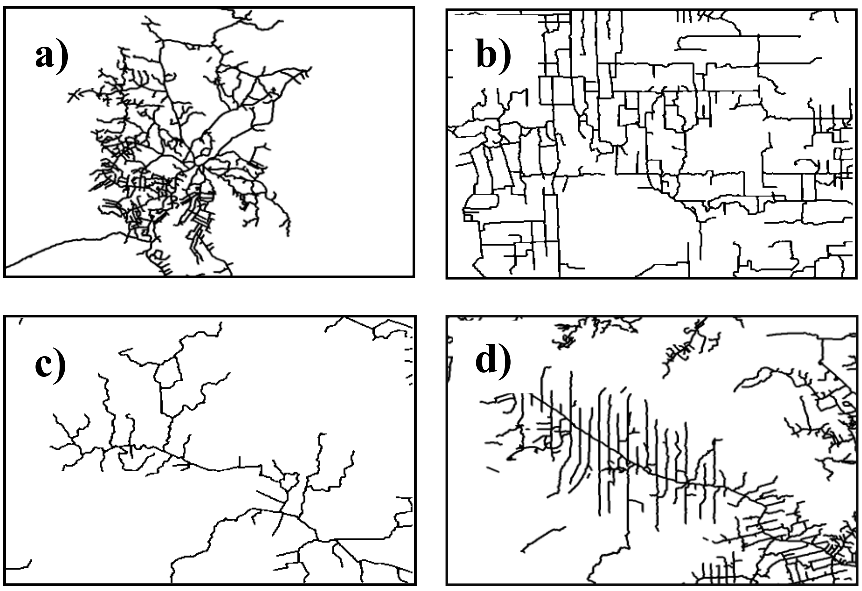


Figure 2.2. Examples of main road network patterns a) dendritic, b) rectangular/grid, c) radial, d) fishbone. Images taken from a map of the Amazon road network, provided by IMAZON.

2.3. What determines road building?

The underlying drivers of deforestation are very similar to the underlying factors that determine the development of road infrastructure. Economics, policy, technology, demographic and cultural factors all influence the rate, location and extent of road building. The economic climate has a clear influence as it can determine how much capital is available for investment in infrastructure. Markets also determine how much capital is allocated to different developments, for example if demand for timber increases (and the market value increases) there is likely to be an increase in investment in the road network in order to increase timber extraction. Government policies greatly influence investment in roads, with the government likely to provide subsidies to road builders to an area or indeed directly invest in the network by building federal roads. This process was exemplified by the drive to colonise new areas in the Brazilian Amazon in the 1970's (Carvalho *et al.* 2002). Technological advancements influence the cost effectiveness of investments in the road network. Demographics play a role because as a population increases a better infrastructure is required to provide for the population. Cultural issues include attitudes, values and beliefs towards roads that might influence their spatial patterning. For example, very few people would dispute a road being built through unused waste land, but many would be against a road going through a nature reserve. These forces work at a large scale. On a smaller scale the exact location of a road depends on two main considerations: (1) where the road should go, i.e. where does it start from and where is its destination, and (2) constraints on the alignment of the road that impact its feasibility and/or cost, such as rivers, mountains and human land uses.

Some roads have a very definite start and end point that is decided before construction starts. These are termed destination determinate roads, and usually occur when a road is being built for a very specific purpose. For example, a road constructed to connect a new housing development to a major highway has a clearly defined start and end point and is a good example of a destination determinate road. By contrast, roads developed to aid in resource extraction, such as logging roads, may not always have a definitive end point. When a timber company wishes to access timber in a new location, they may simply cut a pathway (which will become a road) into the forest, continuing to add and extend pathways until the desired timber resource has all been accessed and extracted. In this case, there is no pre-determined destination, although there is a pre-determined area that the road network must encompass. These roads are termed destination indeterminate roads. Of course, not all logging roads are destination indeterminate, and many networks of logging roads will be a combination of the two. For example, a timber company will develop a destination determinate road to a known stand of timber for which they have logging rights, but once in that stand additional roads to extract that timber will likely be destination indeterminate.

Road alignment is the location of the road in relation to the surroundings, describing the spatial layout of the road. Roads are built in three-dimensions, so road alignment includes both a horizontal (forward and backward, left and right) and a vertical (up and down) aspect. The alignment is dependent on a range of factors that can either constrain or facilitate the laying of the road (Koorey 2009), and these need to be taken into consideration as parameters when modelling road networks. Constraints on alignment fall broadly into five categories:

1. Topography. It is harder to build in hilly or mountainous areas because of steep inclines and because there is a high chance that earthwork, such as cutting or banking of the ground will need to be carried out. Topography is considered the main influence on where a road will be built (Arima *et al.* 2005).
2. Existing development and land use. It usually takes longer and costs more to build in or through pre-existing developments such as urban areas. This is because constructions in urban areas need more planning to minimise disruption and because land prices are much higher. Existing developments act as an alignment constraint because it is often not feasible to build a new road over an existing development. Protected areas such as nature reserves and archaeological sites also pose a constraint on road alignment because of their protected status.
3. Hydrological features. Rivers and lakes form obvious constraints on road alignment. Rivers are not absolute barriers in the way that lakes can be, but do add to the costs of road building when bridges are required. Rivers may also form part of a transport network but for road building purposes rivers are considered a semi-barrier because they can affect the alignment of a new road.
4. Ground conditions. Ground conditions include factors such as substrate (sand vs. clay), compaction (soft vs. hard) and drainage (swamp vs. dry land). These present obvious constraints on road alignment with it being easier to build on solid dry ground than on wet swampy ground.
5. Curvature. Curvature is the deviance a road takes from a straight line and is particularly important in relation to safety and ease of road use. Some road accidents that occur are due to faults in the road alignment (Roh *et al.* 2003). Roads of different widths, on different inclines and used for different purposes have different legal

maximum curvatures, and appropriate curvature is known to reduce accidents in urban areas (Haynes *et al.* 2007, Haynes *et al.* 2008).

2.4. Ecological effects of roads

Since the 1900's, global road networks have been expanding at a rapid rate (Forman *et al.* 2003). Road networks were first studied by 'transportation geographers' (Coffin 2007) whose main concerns were structural network properties, economics and development. From this early work a number of quantitative methods were developed for the study of networks (Coffin 2007), yet little attention was given to the environmental impacts that the road networks had. One impact however, namely road mortality, has been at the forefront of research on road effects from as early as 1935 (Stoner 1935).

By the 1970's, research on the effects of roads on 'wildlife' began to emerge in earnest, with work centring on three main topics: 1) road mortality; 2) roads as barriers; and 3) roads inducing behavioural changes in animals. In the 1980's, the field of landscape ecology began to establish (Wiens *et al.* 2007), and with it came a strong focus on the effects of scale and fragmentation patterns. Given roads are a major force in fragmenting natural habitats, it is unsurprising that in recent years attention has been turned towards the effects of roads on landscapes and ecology, and has even led to the emergence of a new field coined 'Road Ecology' (Forman 1998). This research field focuses on understanding the interactions between road networks and the natural world, and is growing at an exponential rate (Figure 2.3).

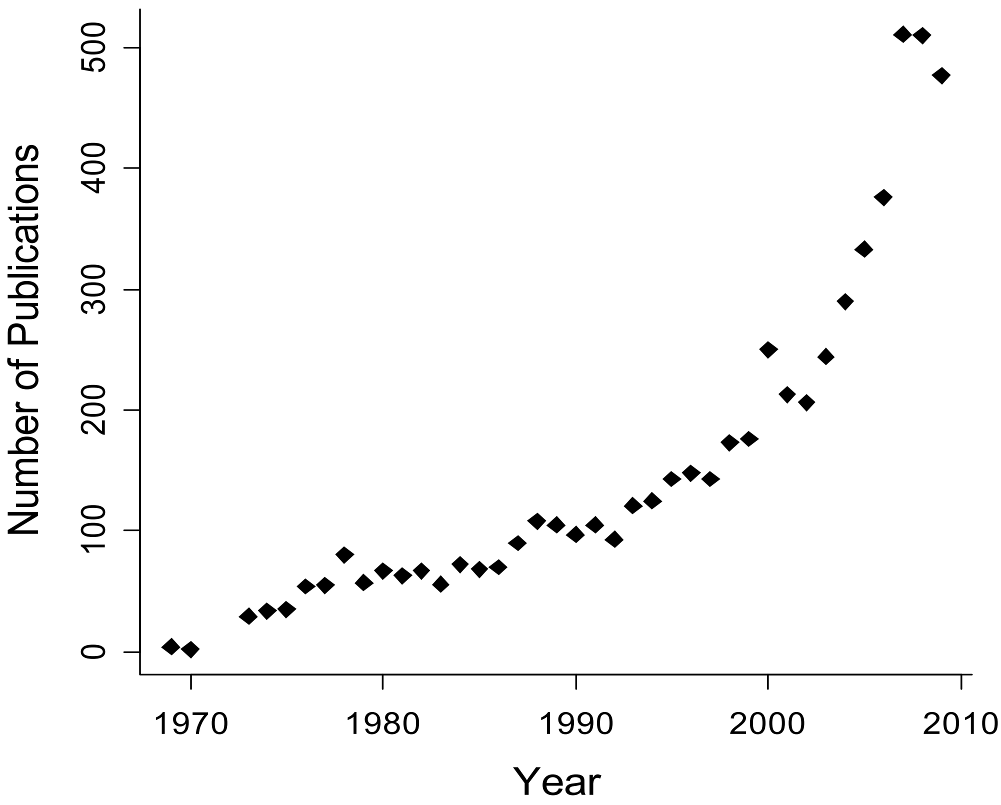


Figure 2.3. The number of scientific papers published each year that use the term ”road ecology”. Data were obtained from the Web of Knowledge literature database, searched on 30 March 2010, which returned a total of 5,902 articles.

Roads have many and varied ecological effects, many of which are difficult to categorise into discrete themes. Most often, ecological effects fall into multiple categories or there are associated knock on effects and links between categories (Figure 2.4). Some road effects act at large scales over long periods of time, such as traffic pollution that has long term implications for climate, whereas other effects have more localised and short term impacts such as isolated incidences of road mortality. The magnitude of the ecological impacts are determined by a range of factors, including: (1) the scale of the road or road network (physical size); (2) the level of use (traffic flow) with a busy road likely to have a larger impact than a quiet one; and (3) the time of road use which will moderate the magnitude and target of the impacts. For instance, if traffic is highest at night, nocturnal animals may suffer more than diurnal animals. Geographic location also plays a role in determining the ecological impacts of roads. For example, pollutants from de-icers will be an issue in cold regions but irrelevant in hot regions, creating a latitudinal gradient in the nature of road impacts. At a smaller spatial scale, roads on the side of hills vs those on flat land will have different impacts on hydrology and erosion rates.

Impacts from roads often occur beyond the immediate vicinity of the roads themselves, impacting much greater spatial areas than might be expected from the size of the network alone. The extent and direction in which road impacts are transmitted beyond the physical boundary of a road can be highly variable. Some effects occur far from the road itself, such as the quarrying and manufacture of road-building materials, whereas impacts such as road mortality are tightly constrained to the location of the road itself. In between these two extremes lie buffer effects, which shadow the spatial pattern of road networks but extend beyond the road itself, such as light pollution from road lights. Buffer effects, also called the 'road-effect zone', extend variable distances from the road edge depending on the specific effect (Forman & Deblinger 2000, Coffin 2007). Some road effect zones are directional, with hydrological changes and erosion patterns having knock-on effects that are transmitted downstream and downhill, but probably not upstream or uphill. Roads also cause changes to abiotic processes, which in turn can influence biotic responses. For example, roads through forests can create open edges that have increased exposure to the sun and altered microclimates, which in turn can cause a shift in animal and plant distributions.

Some ecological impacts are incremental and cumulative, with the impact growing as the road network grows or as the road is in operation for longer. For example, the spatial area impacted by edge effects will increase as the road network grows over time. By contrast, some effects may be felt only in the short term, such as a pulse of sediment into streams during the road building process. Importantly, road effects can change with the life of the road; the effects during the construction, operation, maintenance and de-commissioning or abandonment phases of a road will all differ. In a similar vein, some effects are incidental, arising as a result of people using roads for purposes other than what they were initially constructed for. For example in Africa, roads built to extract timber for the forestry industry are used by people to gain access to forest for hunting bush meat (Wilkie *et al.* 2000).

The variety of effects and the fact that roads impact more of an ecosystem than would be indicated by their physical footprint (much like a keystone species impacts its environment more than expected) means that roads could be considered ‘keystone landscape elements’ (McGarigal *et al.* 2001). The rest of this section is an overview of the breadth and extent of road effects. It serves to illustrate why roads are important in understanding environmental changes and why it is important to model them.

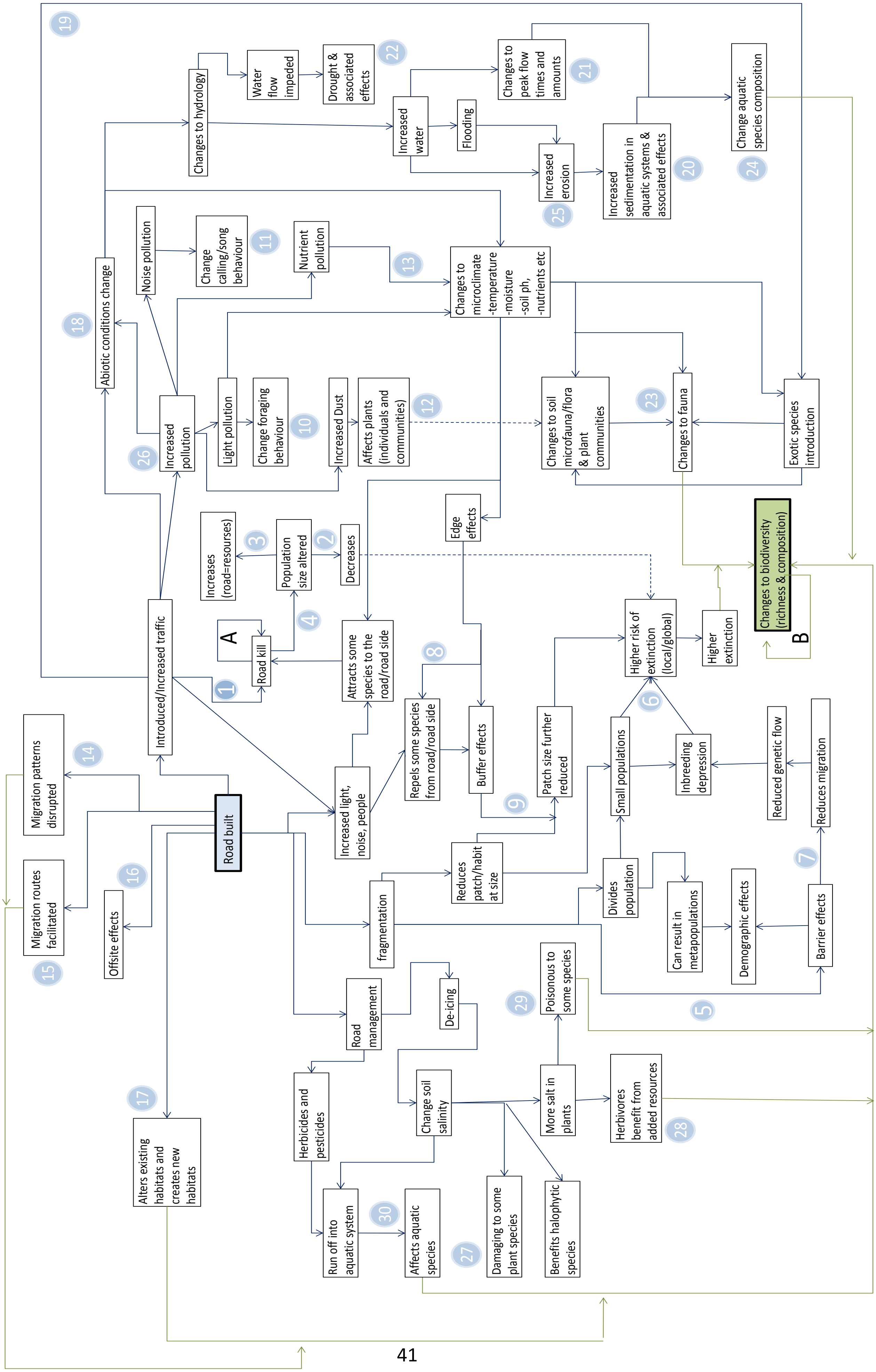

Road built
Introduced/Increased traffic
Road kill
A
Population size altered
Increases (road=resourses)
Decreases
Attracts some species to the road/road side
Increased light, noise, people
Repels some species from road/road side
Buffer effects
Edge effects
Increased pollution
Light pollution
Change foraging behaviour
Noise pollution
Change calling/song behaviour
Nutrient pollution
Increased Dust
Affects plants (individuals and communities)
Abiotic conditions change
Changes to hydrology
Water flow impeded
Drought & associated effects
Increased water
Flooding
Changes to peak flow times and amounts
Increased erosion
Increased sedimentation in aquatic systems & associated effects
Change aquatic species composition
Changes to microclimate -temperature -moisture -soil ph, -nutrients etc
Changes to soil microfauna/flora & plant communities
Changes to fauna
Exotic species introduction
Changes to biodiversity (richness & composition)
B
Higher risk of extinction (local/global)
Higher extinction
Patch size further reduced
Small populations
Inbreeding depression
Reduced genetic flow
Reduces migration
Reduces patch/habit at size
fragmentation
Divides population
Can result in metapopulations
Demographic effects
Barrier effects
Migration patterns disrupted
Migration routes facilitated
Offsite effects
Alters existing habitats and creates new habitats
Road management
De-icing
Herbicides and pesticides
Change soil salinity
Poisonous to some species
More salt in plants
Herbivores benefit from added resources
Damaging to some plant species
Benefits halophytic species
Run off into aquatic system
Affects aquatic species
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30

Figure 2.4. Major ecological effects of roads, how they are linked and how they affect biodiversity. Loops, (A) Road kill attracts scavengers resulting in more road kill, (B) changes in species richness and composition has knock on effects via food webs and species interactions that result in further diversity/composition changes, these changes may be subject to time lags (Findlay & Bourdages 2000). **Figure citations**; (1) Stoner 1935, Forman & Alexander 1998, (2) Kociolek *et al.* 2011, (3) Kristan *et al.* 2004, (4) Fahrig &Rytwinski 2009, (5) Keller & Largiader 2003, (6) Caughley 1994, (7) McGregor *et al.* 2008, (8) Lehman *et al.* 2006, (9) Coffin 2007, (10) Rydell 1992, (11) Brumm 2004, Slabbekoon & Ripmeester 2008, (12) Farmer 1993, (13) Angold 1997, (14) Sawyer *et al.* 2005, (15) Trombulak & Frissell 2000, (16) Spellerberg 2002, (17) Lugo & Gucinski 2000, (18) Spellerberg 2002, Forman & Alexander 1998, (19) Forman & Alexander 1998, Parendes & Jones *et al.* 2000, (20) Jones *et al.* 2000, (21) Jones *et al.* 2000, (22) Trombulak & Frissell 2000, Laurance *et al.* 2009, (23) Fagan *et al.* 1999, (24) Bain *et al.* 1988, (25) McGarigal *et al.* 2001, (26) Bingal *et al.* 2007, (27) Spellerberg 2002, (28) Laurian *et al.* 2008, (29) Mineau & Brownlee 2005, (30) Sanzo & Hecnar 2006.

2.4.1. Road mortality

Mortality is one of the most obvious, and one of the first road effects to be studied, with the literature dating from the 1930's (Stoner 1935, Scott 1938). These recordings were primarily concerned with large mammals and tended to present observations of mortality (Stoner 1935, Scott 1938, Pickles 1942). Even today the literature is primarily concerned with vertebrates (Shyama Prasad Rao & Saptha Girish 2007). More recently, the effects of road mortality on populations and demography have moved the field from empirical observations to predictive modelling (Row *et al.* 2007, Clevenger *et al.* 2003, Jaeger *et al.* 2005, Jaarsma *et al.* 2006, Ortowski 2008, Glista *et al.* 2009, Roger *et al.* 2010). Road mortality directly reduces population size, however, for most species, the loss of individuals through road mortality is not a significant determinant of population survival (Adams & Geis 1983, Munguira & Thomas 1992, Forman & Alexander 1998, Hels & Buchwald 2001, Seiler *et al.* 2004, Orlowski & Nowak 2006, Munro *et al.* 2012). This, however, depends on the frequency of deaths and specific species traits (Hodson 1962, Fahrig & Grez 1996, Carr & Fahrig 2001, Barthelmess & Brooks 2010, Caceres 2011, Rytwinski & Fahrig 2011). Species with low population densities and/or low reproductive rates will be more severely impacted than species with high reproductive rates and population sizes, because the loss of each individual has a higher impact on the overall population. Florida Scrub Jay (*Aphelocoma coerulescens*) (Mumme *et al.* 2000), Audubon's Crested Caracara (*Polyborus plancus*), the Hawaiian Goose (*Branta sandvicensis*) (Kociolek *et al.* 2011), Barn owl (*Tyto alba*) (Fajardo 2001), Little owl (*Athene noctua*) (Hernandez 1988), Spotted turtle (*Clemmys guttata*) and Blanding turtle (*Emydoidea blandingii*) (Beaudry *et al.* 2008) are examples of species that do suffer population declines as a result of road related mortality.

Road mortality events can be beneficial to some species. Species that eat road kill, such as ravens and vultures (Kristan *et al.* 2004, Kelly *et al.* 2007), and which have the capacity to avoid traffic, show an increase in abundance and thus benefit from road mortality events (Fahrig & Rytwinski 2009). For example, carrion eaters take advantage of road kill, which may be seen as a diet subsidy. Common ravens (*Corvus corax*) that have a greater content of road kill in their diet have greater fledgling success (Kristan *et al.* 2004). Further, a survey of the foraging behaviour of 1,947 ravens found that 21 % of all feeding and foraging behaviour was related to road kill events (Dean & Millton 2003).

2.4.2. Fragmentation

Roads can fragment habitats and act as barriers to dispersal; they present a disjunction in habitat that that many animals avoid crossing. This impact may be magnified by road mortality (forming an 'absorbing' barrier), but in most cases the avoidance is behavioural with species avoiding the road itself. For some species, roads are 'absolute' barriers that are never crossed (Keller & Largiader 2003), although for many species the road forms a semi-permeable barrier that individuals actively avoid crossing. Crossing avoidance has been observed across many taxa including, mammals (Richardson *et al.* 1997, Dyer *et al.* 2002, Rico et al 2007, McGregor *et al.* 2008), birds (Laurance *et al.* 2009, Tremblay & St Clair 2009), amphibians (Marsh *et al.* 2005), reptiles (Shepard *et al.* 2008) and invertebrates (Keller & Largiader 2003, Bhattacharya *et al.* 2003).

Road avoidance behaviour affects species' distribution resulting in, range shifts, range restrictions and changes in habitat use, by acting as barriers or buffers. The Moustached monkey (*Cercopithecus cephus)*, Grey-cheeked monkey (*Lophocebus albigena)*, Agile mangabey (*Cercocebus agilis)*, Amur tigers (*Panthera tigris altaica*), Elephants (*Loxodanta africana cyclotis),* Red duikers (*Cephalophus* spp.), Oven birds (*Seiurus aurocapillus*) and Woodland salamanders (eg *Plethodon metcalfi*) are all encountered significantly less near roads (Ortega & Capen 1999, Kerley *et al.* 2002, Potvin *et al.* 2005, Whittington *et al.* 2005, Blom *et al.* 2005, Semlitsch *et al.* 2007). In some cases this is a result of the road itself, for example with woodland salamanders that avoid logging roads even once they have been abandoned (Semlitsch *et al.* 2007). In other cases avoidance occurs as a result of the positive relationship between roads and other human pressures such as hunting for example with wolves (Whittington *et al.* 2005) and in some cases avoidance is because roads reduce habitat quality (Ortega & Capen 1999).

The level of road avoidance, and therefore the level of impact that road induced fragmentation might have on a population, is determined by the interaction of species traits with road characteristics. Species with large territories, species that are easily disturbed by light and noise, and species that use habitat cover for movement, are more impacted by road fragmentation. A small or narrow road is less of a barrier than a large or wide road, as evidenced by studies on carabid beetles (Yamada *et al.* 2010), understory birds of the Amazon (Goosem 2007, Laurance *et al.* 2009), small mammals (Goosem 2007, Rico et al 2007) and obligate arboreal vertebrates (Gossem 2007). The difference in impact can be large, with data from small rodents crossing forest roads showing that movement rates across roads were reduced by 67-90% across narrow clearings and by 90-100% across wide clearings (Laurance *et al.* 2009). Concomitant with road width is traffic density which also influences the permeability of a road barrier. For example, Chruszcz *et al.* (2003) found that Grizzly bears (*Ursus arctos)* are more likely to cross roads with low traffic density. For other species, traffic density has no effect because animals are avoiding the road itself, because it is an open, vulnerable location, rather than avoiding the various emissions from vehicles, such as light and noise (Rico *et al.* 2007, McGregor et al 2008). The degree of road avoidance can be further modulated by intra-specific trait variation. For example, female panthers (*Puma concolor*) avoid road crossings, but male panthers readily cross roads (Kerley *et al.* 2002).

For some species the presence of individual roads is not a deterrent, but the overall density of the road network is a key determinate of habitat selection. Grizzly bears (*Ursus arctos)* (McLellan & Shackleton 1988, Mace *et al.* 1996), elk (*Cervus elaphus roosevelti*) (Witmer & deCalesta 1985), wolves (*Canis lupus*) (Thiel 1985, Mech *et al.* 1988, Whittington *et al.* 2005, Potvin *et al.* 2005) and amphibians (Vos & Chardon 1998, Eigenbrod *et al.* 2008), all preferentially locate in areas of lower road densities. This could be because areas of low road density areas experience less human impact and disturbance. In fact, road density was found to be more important than forest cover for habitat choice in three frog species; *Bufo americanus, Rana pipiens* and *Hyla versicolor* (Eigenbrod *et al.* 2008).

Road networks gradually build up over time, and as a result shifts in species' ranges are considered cumulative effects (Whittington *et al.* 2005). A region that previously did not have a road density high enough to force certain species to avoid it can find those species progressively excluded as road density increases. However, despite many animals exhibiting a negative relationship with road density, some animals benefit from high road densities. White-footed mice (*Peromyscus leucopus)* avoid road crossing but the negative impacts of this seem to be outweighed by a positive effect on their abundance near roads. Rytwinski & Fahrig (2007) suggest two possible reasons: (1) roads are positively correlated with an undetermined component of habitat quality; or (2) roads negatively impact White-footed mice predators.

Fragmentation results in smaller suitable habitat patches, which inevitability have lower carrying capacities than large habitat patches, resulting in reduced population sizes. Further, roads act as barriers to movement, resulting in these smaller populations being isolated (if the road is an absolute barrier) or establishing metapopulations (if there is some movement across the roads). Either way, small populations are at greater risk of extinction as a result of stochastic demographic and environmental shifts (Caughley 1994). Roads reduce re-colonisation of empty habitats by limiting immigration (McGregor *et al.* 2008), and that increased isolation reduces gene flow, which combined with small population sizes, can result in inbreeding depression.

As roads fragment habitat, they create edges leading to edge effects; defined as the ecological effects arising as a result of interactions between adjacent habitats that are separated by a transition zone that is usually abrupt (Murcia 1995). Road building typically creates new edge effects because the road presents a new environment that is juxtaposed with, or more usually passes through, an existing habitat. Roads induce drastic abiotic edge effects along their borders, which include changes to the microclimate; light levels generally increase, air and soil temperature and moisture change because of increased exposure, soil pH and nutrient levels change because of roadside management and the introduction of a road surface which is chemically different to the native habitat (Delgado et al 2007, Honu & Gibson 2006, Gehlhausen *et al.* 2000). Changes to abiotic conditions can have knock on effects, for example forest edges along roads tend to be drier than forest areas with no roads, and as such they are more prone to fire (Cochrane & Laurance 2002).

Changes in the abundance of species (Marsh & Beckman 2004, Donovan et al 1997, Lehman et al 2006), distribution of species (Lehman *et al.* 2006, Baldi & Kisbenedek 1999) and introduction of alien species (Honu & Gibson 2006) can all occur as a direct result of edges being present and changes in abiotic conditions near the edge. The spatial scale of road induced edge effects increase over time as the network grows. Further, road edges are different to naturally formed edges because eventually they 'box in' a patch (i.e. they form around the patch perimeter). This boxing in is particularly problematic as the network grows because patches become smaller with the effective patch size being further reduced by buffer effects. As a result of the negative effects of road induced fragmentation, strategies attempting to mitigate these effects have been made, including, over- and under-passes, corridors, canopy bridges and areal stepping stones (Colcheo *et al.* 2011, Taylor & Goldingay 2011, Goosem 2012, Lesbarreres & Fahrig 2012).

2.4.3. Behaviour

Roads fragment habitats and reduce animal movement, alter species ranges and the habitat selection patterns of individuals. All of these effects are mediated by animal behaviour or, most specifically, road avoidance behaviour. Conversely, species that benefit from roads might be attracted to habitats near or containing a high density of roads. The distributions of the Turkey vulture (*Cathartes aura*) and Black vulture (*Coragyps atratus*) are influenced by the distribution of carrion such as road kill (Kelly *et al.* 2007). Raptors (Accipitridae and Falconidae) are attracted to roads because, although they tend not to feed on road kill, they are attracted to the productive road-side verges that are often good habitats for small prey (Dean & Millton 2003). Kangaroo rats (*Dipodomys ordii*) benefit from easier digging, dust bathing and higher seed banks found along road edges (Stapp & Lindquist 2007). Basking lizards and snakes take advantage of the increased temperature of tarmac (Vijayakumar *et al.* 2001). Herbivores may also be attracted to road side habitats where vegetation has higher concentrations of salts and nutrients, usually from de-icers, fertilizers and other road side pollutants. For example, moose (*Alces alces*) generally avoid roads but are attracted to roadside vegetation along roads that are de-iced with salt, because the plants contain a higher level of sodium, which can be a limited resource (Laurian *et al.* 2008).

Roads may also alter migratory patterns, with some animals avoiding routes near roads but others making use of the easier path that a road offers as a movement corridor. For example, Mule (*Odocoileus hemionus*) and Pronghorn deer (*Antilocapra Americana)* experience bottlenecks in their migration routes as a result of roads and housing developments (Sawyer *et al.* 2005) and it is thought that the building of a new road in the Serengeti will result in disruption to Wildebeest migration (Dobson *et al.* 2010). Conversely caribou (*Rangifer tarandus*) utilise cleared winter roads in the direction of their normal migratory pattern (Trombulak & Frissell 2000). Cane toads (*Bufo marinus*) and wolves have also been shown to utilise roads as movement corridors (Brown *et al.* 2006, Forman *et al.* 2003). Species that utilise roads for movement may gain access to previously unoccupied habitats and thus expand their range.

Road lights, light from passing vehicles and the way light 'interacts' with the road, can all deter or attract animals, causing changes to normal behaviour. Light pollution is one of the most rapidly increasing changes to the environment (Cinzano *et al.* 2001). Approximately two thirds of the world's population and 99 % of the European Union population are in areas where the night sky is above the threshold for light polluted status. Large areas of natural and semi-natural areas are exposed to light pollution from nearby urban areas and roads (Santos *et al.* 2010). Illumination from roads could be seen as more invasive than that from urban areas because roads and their associated light penetrate into natural habitats. Artificial illumination from roads parallel to beaches (and beach front developments) causes disorientation in baby sea turtles, who orient themselves towards the sea by using patterns of light reflected from the sea (making it brighter) and absorbed by the beach/vegetation behind the beach (Tuxbury & Salmon 2005, Santos *et al.* 2010). Fledglings of sea birds also experience disorientation from artificial light as they attempt to reach the sea for the first time (LeCorre *et al.* 2002), and many die as a result of injuries, starvation or predation, because of failure to reach the water quickly (Santos *et al.* 2010).

Light pollution from roads and roadsides can also have positive impacts on some species. Nocturnal predators experience greater visibility for hunting and others can feed upon concentrations of insects that are attracted to lights (Rydell 1992). Large numbers of congregated insects provide an ideal foraging location for insectivorous bats (providing they are able to avoid traffic). A study by Rydell (1992) showed that the gross energetic intake of *E. nilssonii* foraging around road lights was more than twice as high as those foraging in wood lands (0.5kJ/min compared to 0.2kJ/min) as a result of lights attracting energy rich moths (as opposed to flies in woodlands). Diurnal animals may extend their daily activity as a result of the extra light.

Road surfaces can interact with natural light, mimicking cues and signals that some insects rely on for normal mating behaviour. Mating and egg-laying mayflies (Ephemeroptera) are attracted to asphalt roads because reflected light is strongly, horizontally polarised, which makes it appear like a water surface to insects that seek water based on polarotaxis (partial and horizontal polarisation of reflected light). Mating mayflies are further attracted to roads because of their elongated shape (much like a stream) and because there is no overhanging vegetation (a prerequisite for mating). This change in reproductive behaviour (mating and laying eggs on asphalt instead of water) is damaging to mayfly populations because eggs laid on asphalt perish (Kriska *et al.* 1998, Kriska *et al.* 2007). Most insects whose larvae develop in freshwater use polarotaxis to locate water sources (Kriska *et al.* 2009), including dragonflies (Odonata) and tabanid flies (Tabanidae), suggesting that many insects' reproductive behaviour can be interfered with by asphalt covered roads.

Roads are a source of ambient noise in the environment, with the level of noise pollution determined by the flow and weight of traffic. Noise acts as a strong deterrent to many species, keeping them away from the road vicinity. For example, some species of foraging bats will avoid areas with roads in favour of silent areas (Schaub *et al.* 2008). Light and noise pollution causes changes in animal foraging behaviour (Slabbekoon & Ripmeester 2008). Additional light can affect foraging behaviour, with bats that previously scanned for food over large areas now limiting their foraging area to well lit roads. A study by Santos *et al.* (2010) found that visually foraging wading birds increased foraging effort in artificially illuminated areas, and that waders that used a mixture of visual and tactile foraging favoured more effective visual foraging style in light polluted areas. These shifts in foraging behaviour increased prey intake rate by an average of 83 %, an obviously positive effect of light pollution. Noise pollution leads to some animals devoting more time than usual scanning for predators in areas of elevated noise, such as near a busy road. Chaffinches, *Fringilla coelebs*, were found to spend less time foraging during artificially increased noise levels (Quinn *et al.* 2006) in order to 'look' for predators because auditory stimuli detection is reduced.

Animals such as birds and amphibians use calls and songs to attract mates and stake territory (Bee & Swanson 2007). Traffic noise from roads interferes with these acoustic signals and has led to changes in singing behaviour. Birds have been shown to increase song amplitude (volume), known as the Lombard effect, to compete with traffic noise (Brumm 2004) both in the field and in experiments using white-noise. Examples of birds that the Lombard effect has been shown in include, zebra finches *Taeniopygia guttata* (Cynx *et al.* 1998), budgerigars *Melopsittacus undulatus* (Manabe *et al.* 1998), blue-throated hummingbirds *Lampornis clemenciae* (Pytte *et al.* 2003), Nightingales *Luscinia megarhychos* (Brumm 2004) and domestic fowl (chickens) *Gallus gallus* (Brumm *et al.* 2009). An alternative to changing call amplitude is to change call frequency (pitch). Birds will generally increase the frequency of their signalling to avoid masking by traffic (which is usually low frequency) (Slabbekoon & Ripmeester 2008, Halfwerk & Slabbekoon 2009, Parris & Schneider 2009). Finally, many species alter their temporal pattern of acoustic signalling to avoid masking and interference from other species' calls (Warren *et al.* 2006). Given this, a temporal shift in signing activity in bird species, competing with traffic, is not unexpected. Such a shift usually changes diurnal singing patterns to avoid peak traffic. One such example is the European robin, *Erithacus rubecula,* (Fuller *et al.* 2007) that has been found to sing nocturnally in areas of high traffic.

Acoustic behavioural responses have also been seen in monkeys and frogs. Brumm *et al.* (2004) played white-noise to common marmosets (*Callithrix jacchus*) and found that in addition to increasing amplitude they also increased the duration of calls (although not studied with traffic, a similar response can arguably be expected from marmosets near roads). However not all animals are capable of changing their behaviour to compensate for road presence. For instance, Tree frogs (*Hyla arborea*) are not able to adjust their call frequency or temporal structures in response to traffic-noise and are therefore unable to transmit information to each other, reducing reproductive success (Lengagne 2008).

Responses of animals to noise pollution from roads are an issue of increasing concern for conservation and animal behaviour biologists (Warren *et al.* 2006). Behavioural responses may be short-term phenotypically plastic responses, long-term phenotypically plastic responses (e.g. song learning) or may be evolutionary responses under natural selection (Warren *et al.* 2006), thus roads are capable of driving evolution as well as of shaping the landscape.

2.4.4. Habitat structure

Roads alter the structure of the landscape and associated habitats (McGarigal *et al.* 2001, Saunders *et al.* 2002). As soon as a road is laid habitat is destroyed and the remaining habitat is fragmented, during road operation edge effects reduce the suitability/quality of habitats for interior species. In one study it was found that mean patch size and core habitat area declined by 40 % and 25 % (respectively) as a result of logging road development over 40 years (McGarigal *et al.* 2001). On the other hand, roads provide new habitats, for example bridges provide new nesting sites for birds (Forman 1998, Kociolek *et al.* 2011) and road verges form new succession sites (Forman *et al.* 2003), thus the age of a road influences the community structure present (Spooner & Smallbone 2009). Alternatively, road verges can be planted and maintained, altering the original local diversity and structure of plant communities. The specifics of how a road edge/verge is managed can modify the effect the road has on biodiversity. For example, roads that have trees within 20 meters have lower owl mortality than those with trees more than 20 meters away, and roads with perches (trees/shrubs/hedgerows) taller than two meters experience less owl mortality than those with perches shorter than two metres (Hernandez 1988, Orlowski 2008). Small scale habitat structure is also altered by road presence, such as lower leaf litter depth near road edges that is possibly due to increased exposure of edges to wind (Haskell 2000). Changes to habitat structure leads to many abiotic changes, for example microclimates, hydrology, erosion rates and biogeochemical cycles are altered as a result of landscape and habitat structure change.

2.4.5. Microclimate

The microclimate surrounding a road differs from the microclimate of the surrounding natural environment for two main reasons: (1) the road surface has a different albedo to the surrounding habitat; and (2) roads are exposed and they expose the edges of the surrounding habitat. These two factors combine and result in differing microclimatic dynamics depending on the road and what the surrounding habitat is. For instance a small, quiet road passing through low grassland is likely to have less effect on the microclimate than a large, busy road passing through dense forest. Roads typically have a lower albedo than natural habitats and so are generally warmer than surrounding areas. This is taken advantage of by basking reptiles (Vijayakumar *et al.* 2001) and birds that rest on road surfaces, reducing their metabolic costs (Kociolek *et al.* 2011). By creating exposed edges, the area immediately surrounding a road has a higher temperature than further away from a road (especially in forested habitats). However, edges are more exposed to wind (natural or generated by traffic) which can reduce edge temperatures. By altering exposure, roads also change light levels and humidity. Changes in soil moisture can be attributed to roads in several ways; changes in runoff rates as a result of the impervious nature of road surfaces increase moisture levels, while increased exposure along a road edge will result in drier soils. Roads do alter the microclimate but their effects are varied and depend on the changes and interactions between pre-existing habitat, road characteristics, and specific changes in exposure and light conditions.

2.4.6. Hydrology

Roads alter the hydrology of landscape in many ways, primarily through forming a hard, compacted surface that alters the flow of water run-off that can cause major changes in terrestrial (Young 1994) and aquatic systems (Forman & Alexander 1998, Jones *et al.* 2000, Coffin 2007). Changes in run-off regimes can lead to flooding, with roads increasing the amount of water reaching a stream system (Spellerberg 2002, Forman *et al.* 2003). Roads increase the peak flow of streams and rivers by increasing the amount and rate at which water is introduced via run-off (Jones *et al.* 2000). If the roads have a drainage system that connects to the waterways, then the road network extends the drainage basin of the stream system (Forman & Alexander 1998). Flowing water shapes landscapes via streams and rivers; roads generally result in more water runoff and consequently faster flowing water, faster flow is stronger flow and results in faster changes to the landscape. For example, over time river bends become deeper as the faster flowing water on the outside of the bend cuts into the land and the slower flowing water on the inside deposits sediment.

Faster water flow can alter more than just the shape of the waterway. Aquatic species are adapted to certain flow rates and regimes (Bain *et al.* 1988), and by altering these conditions roads can lead to changes in species composition of the waterway. Species that are adapted to survive in slow moving water may not be able to cope with the increase in flow rate as a result of increased run-off, so stream communities will move from slow water adapted species to those that can cope with increased flow. Also, faster flowing streams have reduced community complexity compared with slow flow stream systems (Bain *et al.* 1988).

Roads increase the natural instability of montane habitats (Young 1994), with an increased frequency of landslides observed in steep-forested landscapes with roads compared with equivalent landscapes with no roads (Jones *et al.* 2000). Increased run-off from roads results in more erosion (McGarigal *et al.* 2001) which, coupled with an increase in landslides, results in increases in the 'debris flow' of stream networks and thus higher deposition of sediment into waterways. Sediment clouds the water and thus changes suitability of the system for many aquatic species.

Increased run-off removes topsoil and reduces the fertility of areas in the run-off path, potentially reducing productivity. Increased runoff and associated erosion introduces an increased amount of chemical pollutants (heavy metals and nutrients), leached from the land, into coastal and inland aquatic systems, which inevitably has knock on effects in these systems (Davidson *et al.* 2010).

2.4.7. Pollution

The presence of a road is inevitably associated with vehicular traffic, which by its nature, introduces chemical and physical pollutants into the environment. Light, noise (discussed above), dust (particular pollutants), chemicals (de-icers and herbicides), metals (lead, nickel, zinc) and gases (carbon dioxide, sulphur, nitrous oxides, volatile organic compounds (VOC's), polycyclic aromatic hydrocarbons (PAH's) are all released into habitats surrounding roads (Bignal *et al.* 2007). Although most pollutants are introduced to the environment via combustion reactions in vehicles, pollutants may also come from road construction or the road surface itself, management regimes or from spillages (e.g. oil/petrol).

Some chemicals released into the atmosphere add to climate change and those that enter habitats alter the chemical composition of soil and waterways, potentially affecting the local fauna and flora. The range of chemical pollutants is large and there is a vast variation in the level of study dedicated to each; with some pollutants being comprehensively investigated (e.g. nitrous oxides) and others hardly studied at all (e.g. PAH's (Spellerberg 2002)).

Metals introduced to the environment by vehicle exhausts have been extensively studied (Spellerberg 2002), although little is known about the ecological impacts of most metal pollutants (Bingal *et al.* 2007). Lead (Pb) is one of the most extensively studied metal pollutants. Lead was previously used in petrol in order to increase octane number (Majdi & Persson 1989) and for 'anti-knock' properties (Storch *et al.* 2003). Although no longer used in most of the developed world, there are still developing regions where leaded petrol is still available and used as fuel. For example in Africa most petrol sold contains between 0.5-0.8 g/L of lead, far exceeding the WHO's guideline of 0.15 g/L (Ebenso & Ologhobo 2008). Lead negatively effects tree root tip growth, with Majdi & Persson (1989) showing that root tips per unit length decreased closer to roadsides, which inevitably negatively impacts the health of trees near roadsides. Snail shells have also been found to be thinner at lead polluted sites, and their tissue contains high levels of pollutants (including lead) which can be passed up the food chain (Ebenso & Ologhobo 2008). However, not all metal pollution has negative impacts, with calcium (Ca) levels increasing near roadsides that are paved with limestone leading to increases in the dry mass of snails and millipides with proximity to these roads because there is more 'acquirable' calcium that can be utilised in shell and exoskeleton growth (Kalisz & Powell 2003). That same calcium, however, can alter the pH of the surrounding soil using limestone or other base-rich materials in acidic areas leads to an increase in soil pH (Kalisz & Powell 2003, Spellerberg 2002). Conversely, introducing acid-rich material results in a pH decrease. A change in pH can be very detrimental to floral communities that are adapted to either basic or acidic environments (Spellerberg 2002).

'Dust' is a particular pollutant consisting of any solid matter that is fine enough to be raised and carried by wind (Farmer 1993). Depending on the material a road is built of, the type of dust raised will vary, with tarmac roads having the least dust and dirt roads the most. Dust may have chemical or physical impacts and the precise nature of these impacts will depend largely on the physical and chemical nature of the road material from where the dust originates. A review by Farmer (1993) found the presence of dust on a leaf surface may smother the leaf reducing photosynthesis. Dust can block stomatal openings and even stop gas exchange, inhibit pollen germination, halt starch production, stimulate leaf necrosis, reduce transpiration, reduce enzyme activity, and ultimately reduce fruit set and increase leaf temperature (which can disrupt biochemical processes).

Nitrogen pollution introduced in the forms of nitrous oxides from vehicle exhausts is beneficial to some plant species. Heather (*Calluna vulgaris*) and other grasses on heathland habitats close to roads experience increased growth in nitrogen polluted areas (Angold 1997). However heather is adapted to low nutrient, acidic soils (Iason & Hester 1993).

When exposed to nitrogen pollution the abundance of heather decreases, despite improved individual growth rates. Grass species' abundance increases, leading to a shift in community composition, with heather and lichens declining and grasses increasing in abundance (Angold 1997). Nitrogen run-off into aquatic systems can lead to eutrophication and algal blooms.

Other pollutants from car exhausts include sulphur dioxide which, although beneficial at low concentrations, generally alters photosynthetic reactions and thereby reduces growth and productivity (Swanepoel *et al.* 2007). Particulate carbon can behave like a fertilizer and alter plant community composition (Bazzaz & Garbutt 1988, Hunt *et al.* 1991), and ethylene which is used in many plant processes including fruit ripening and leaf senescence, causes disruptions to normal plant phenology (Taiz & Zeiger 2006).

Road management regimes such as de-icing, herbicide and pesticide treatments introduce various pollutants into the environment. De-icers increase the salinity of roadside soil which has effects on the surrounding vegetation and, when transported in run-off, causes negative effects in more distant vegetation and aquatic systems. Increased salt can kill many plant species and increases the susceptibility of some tree species to fungal infections (Spellerberg 2002). De-icing regimes facilitate the dispersal of halophytic plant species, often shifting community dominance in favour of salt-loving or tolerating species. De-icers increase the salinity of local plant species which can be beneficial to herbivores for whom salt is a limited resource (Laurian *et al.* 2008) but ingestion of salts by birds can be fatal (Mineau & Brownlee 2005, Kociolek *et al.* 2011). Run-off transported de-icing salts cause decreases in weight and response times of frog larval stages (tadpoles) and in high concentrations cause developmental abnormalities and are fatal (Sanzo & Hecnar 2006). Herbicides and pesticides used to keep roads and lay-bys free from unwanted weeds and pests affect other plants and invertebrates in the vicinity and run-off into waterways.

2.4.8. Cumulative effects on ecological communities

Abiotic, individual and population level effects of roads have repercussions at a community level often affecting, species composition, abundance, community structure and diversity. Community composition is altered by roads through creating succession sites, altering microclimates and introducing alien species. Roads play a role in the spread of alien species including plants (Gelbard & Belnap 2003, Brisson *et al.* 2010), invertebrates (Suarez *et al.* 1998, Dong *et al.* 2008, Cameron & Bayne 2009) and vertebrates (Brown *et al.* 2006). Roads also play a role in the spread of pathogens, both native and alien (Jules *et al.* 2002, Urban 2006, Haemig *et al.* 2008). Invasions are facilitated by roads in a number of ways: (1) roads act as conduits/corridors for alien invasions, thus alien species are often more abundant near road edges (Spellerberg 2002, Watkins *et al.* 2003, Shepard *et al.* 2008,); (2) Roads lead to increased human activity; humans and their vehicles often carry invasive species with them over long distances (Jules *et al.* 2002); (3) Roads increase disturbances and disturbed habitats allow easier establishment of alien species (Hobbs & Huenneke 1992); (4) Road induced changes in abiotic conditions improve the suitability of road edges for alien species (Forman & Alexander 1998), with exposed road edges having higher light levels making edges suitable for aliens (Parendes & Jones 2000). The longer a road is active, the higher the chance of an alien introduction becomes because they are under a higher accumulated pressure of potential introductions than newer roads (Cameron & Bayne 2009). For example, six species of invasive earthworms in Canada are spread via road networks and older roads had a greater number and extent (i.e. are present further from the road boundary) of worms (Cameron & Bayne 2009).

Roads also have negative effects on community diversity, by increasing local extinction rates and/or decreasing recolonisation rates via, restricting movement, edge effects, changing abiotic conditions, introducing aliens and increasing human activity (Findley & Houlahan 1997). Vascular plant, invertebrate, amphibian, reptile and bird species richness has been found to be negatively impacted, and community structure altered, by roads (Findlay & Bourdages 2000, Haskell 2000, Watkins *et al.* 2003, Laurance 2004, Fahrig & Rytwinski 2009). However, the effect of roads on community diversity is subject to time lags and may not be evident for decades after road construction (Findlay & Bourdages 2000).

Changes in species abundance and richness can lead to knock on effects through food webs and species interactions. For example, soil macroinvertebrate abundance and richness is depressed near roads, predators that rely on soil invertebrates may face food shortages and thus population reductions. This is thought to be true for ground foraging birds like the Wood Thrush (*Hylocichla mustelina*), Black and white Warblers (*Mniotilta varia),* and also woodland salamanders (Haskell 2000).

Very few generalisations can be made about the ecological effects of roads on biota; each species is likely to respond in a different way to the myriad of changes that roads bring about to the environment. One thing however may be said, overall, the effects of increasingly extensive road networks are negative. The Brazilian Amazon is a region undergoing widespread road network development. Although all of the discussed road effects may not be applicable to a tropical setting, the vast array of road effects that are applicable mean that it is very important that we understand where new roads are likely to emerge.

2.5. The Brazilian Amazon

Forests cover approximately 30% of the Earth's land surface (Bonan 2008) and tropical rainforests are beyond a doubt the most diverse and productive forests on earth, accounting for 33% of all terrestrial net primary productivity (Bonan 2008). Tropical forests only occupy about 11% of global land surface but are home to more than half the world's species (Moran 1993). They are characterised by high temperatures, where the mean temperature of the coldest month is at least 18 °C (Whitmore 1998), abundant rain fall, annually more than 200 cm (Begon *et al.* 1996) and 12 hrs daylight daily (Whitmore 1998). Rainforests are generally found along the equator between the tropics of Capricorn and Cancer, in Asia, Africa, Australia, Central and South America (Begon *et al.* 1996). The neotropical (American) forests are the most extensive rainforests accounting for approximately half the global total of rainforests (Whitmore 1998). Brazil contains more rainforest than any other country, with the forest divided between two areas, the Atlantic forest, a strip of forest less than 50 km wide on the coastal mountains and the Amazon Basin (Whitmore 1998).

The Amazon is the largest remaining area of tropical forest (Foley *et al.* 2007), containing half of the world's tropical forest biome (Betts *et al.* 2008) and covering an area of nearly 5 million km^2 (Moran 1993). The Amazon rainforest accounts for approximately 10 % of the Earth's terrestrial net primary productivity and biomass (Melillo *et al.* 1996, Malhi & Grace

2000). It is a highly biodiverse system (Dirzo & Raven 2003) housing a quarter of all global biodiversity (Betts *et al.* 2008), and many Amazonian species are endemic or endangered (Da Silva *et al.* 2005, WWF 2010). A survey carried out in Manaus (Amazonia, Brazil) of just 10 km^2 found a total of 105 mammal, 319 bird and 134 amphibians and reptile species (Ghazoul & Sheil 2010). At an Amazon wide scale, an estimate of tree species suggests that the Brazilian Amazon has 11,210 species of trees with a DBH (diameter at breast height) of more than 10cm (Hubbell *et al.* 2008), in addition to over 30,000 other plant species, 3,000 fish, 427 amphibians, 378 reptiles, 1294 birds and 427 mammals (Da Silva *et al.* 2005). These diversity counts do not take into account the myriad of invertebrate, fungi and microorganism species that live in the Amazon. New species are continually being discovered in the Amazon, with 637 plant, 257 fish, 216 amphibian, 55 reptile, 16 bird and 39 mammal species discovered and described between 1999 and 2009 (WWF 2010). Again, these measures do not count invertebrates, fungi and microorganisms although these are thought to number in the thousands of new species (WWF 2010).

2.5.1. Ecosystem services of the Amazon forest

The Brazilian Amazon is home to at least 206 different indigenous people/tribes, who speak at least 170 languages (Ramos 1998). The Amazon offers ecosystem services to native tribes, local people and the global community. These services include climate regulation, carbon storage, forest products, eco-tourism and diversity. The biodiversity in the Amazon contributes to, and affects, several ecosystem services including biomass production, invasion resistance, existence value, cultural services (many cultures attach spiritual or religious values to the forest or its constituent species), genetic diversity and resources, ecosystem stability and ecotourism (MEA 2005).

Forests influence climate in many ways through physical, chemical and biological processes. The interactions between these processes are complex, often non-linear and can dampen or amplify anthropogenic climate change (Bonan 2008). The Amazon plays a crucial role in local, regional and global climate regulation. It forms one of three major convection centres in the tropics and helps to fuel the Hadley and Walker circulations, which in turn determine the location and magnitude of convective rainfall (Foley *et al.* 2007, Betts *et al.* 2008). This influences local and regional weather and climate patterns. Estimates suggest that deforestation by 2050 will reduce rainfall in the region by 12-21% due to less efficient water cycling (Spracklen *et al.* 2012). At a global scale, the rainforest has the potential to alleviate or exacerbate climate change (Betts *et al.* 2008). Tropical forests sequester large amounts of carbon and contain approximately 25 % of the carbon in the terrestrial biosphere (Bonan 2008). Tropical forests also play a vital role in the carbon cycle and in carbon flux (Melillo *et al.* 1996). Deforestation is the second largest contributor of anthropogenic carbon emissions to the atmosphere, after fossil fuels (Hall 2008). In fact deforestation accounts for 80% of Brazil's carbon emissions (Hall 2008). Maintaining the forest stabilises the global climate by acting as a carbon store and sink.

The Amazon is the largest river system on Earth; it provides navigable waterways, drinking water, aquatic habitats, hydroelectricity and is a source of food and income for many people (Foley *et al.* 2007). The forest regulates the Amazon water system by influencing the amount and time of water flow, and by determining the levels of nutrients that reach the water (Foley *et al.* 2007). Evidence suggests that the removal of the forest disrupts the water regime in the Amazon even if precipitation rates remain the same (Sahin & Hall 1996, Costa *et al.* 2003). The Amazon provides a host of forest products that are used and often exploited by people.

The most obvious forest product is timber and there has been a great deal of research on the causes and consequences of logging and its relationship with deforestation (Fearnside 1987, Uhl & Guimaraes Vieira 1989, McGarigal *et al.* 2001, Asner *et al.* 2004a, Kirby *et al.* 2006, de Oliverira Filho & Metzger 2006, Fearnside 2008). Non-timber forest products include rattans and bamboos, resins and rubber, ornamental plants, materials for crafts (seeds, fibres, etc.), medicines and insecticides, and food products (Montagnini & Jordan 2005). Forest products are used for sustenance and trade by various social actors, from local indigenous peoples to large multi-national companies.

The great diversity and complex systems of the rainforests have made them a finely balanced ecosystem that is relatively easily perturbed; O'Neill (1976) demonstrated with models that tropical forests had the lowest rate of recovery after perturbation compared to six other ecosystems (with the exception of tundra which had an even lower rate of recovery). As an ecosystem the Amazon is valuable for its diversity and services, which may easily be lost as it is degraded and destroyed.

2.5.2. Road development in the Amazon

Development in the Amazon today is the consequence of policies and development regimes initiated in the 1960's (Carvalho *et al.* 2002). Pre-1960 there was little economic incentive to invest time or capital in the Amazon region, primarily because most of the region was inaccessible, and there were no local markets or social infrastructure (Andersen & Reis 1997). The initial drive to develop the Amazon came in 1964 under military rule (Carvalho *et al.* 2002). It was perceived that the Amazon was a 'vacuum', with low population densities and no development, and it was feared that the region was vulnerable to encroachment by other countries and to illegal trades such as drugs trafficking (Carvalho *et al.* 2002, Perz *et al.* 2007). Thus in the early 1960's the Brazilian government initiated a development regime to stimulate economic growth in the Amazon and bring it from the frontier into the mainstream economy (Andersen & Reis 1997). Road development was at the centre of the regime, opening the frontier and allowing easy access for subsequent colonisation and development (Carvalho *et al.* 2002). 'Operation Amazonia' began in 1966, followed by the 'National Development Plan (PND)' in 1970 and the second National Development Plan (PND II)' in 1974. The concept was that main highways, such as the Trans-Amazon highway (BR-230) and Cuiaba-Santarem (BR-163) built in the late 1960's and 1970's, would connect chosen regions, facilitating colonisation and settlement. With the aid of financial incentives, these regions were colonised and developed; ranching, agriculture, industry, logging and service sectors were all encouraged. In 1995, 'Brasil em Acao' (Brazil in action), another development initiative, was established, this initiative was updated in 1999 with 'Avanca Brazil' (Advance Brazil). 'Avanca Brazil' proposes investments of US$500 billion for 358 projects, 21 % of which is allocated to infrastructure development including the construction of new roads and paving of existing roads (Carvalho *et al.* 2002). More recently IIRSA; Initiative for the Integration of the Regional Infrastructure of South America, plans to connect the road networks of south America into one large network. The aims of more recent development initiatives mirror those from the past; a desire to integrate the Amazon through colonisation and development of roads, agriculture and industry, while boosting the economy and raising living standards. While primary roads in the region are planned, the majority of road development is in the secondary and tertiary road network, which is often unplanned and carried out by non-state actors.

Various studies have shown that roads do facilitate subsequent colonisation and development (Fearnside 1987, Verissimo *et al.* 1995, Mertens *et al.* 2002, Arima *et al.* 2005, Caldas *et al.* 2010, Laurance *et al.* 2009, Perz *et al.* 2007), even temporary roads used by loggers and miners for resource extraction are often utilised by settlers. For example, timber companies searching for mahogany (*Swietenia macrophyla*) in the 1980's were the main builders of roads in the Amazonian state of Pará (Verissimo *et al.* 1995). Around the same time, gold miners were also building unofficial roads in Pará (Mertens *et al.* 2002). In both cases, settlers took advantage of access to new land, moving along the emerging road networks wherever agriculture or ranching was feasible, causing deforestation as they cleared land (Verissimo *et al.* 1995, Mertens *et al.* 2002). Temporary roads are often later improved and made permanent by various parties; the people who originally built them, local governments or settlers taking advantage of the unpaved roads.

The initiatives of the 1960's/70's were 'successful'; they lead to more than 60,000 km of roads being built in the Amazon Legal region; a sub-region of the Amazon encompassing seven Brazilian states; Acre, Amapa, Amazonas, Para, Rondonia, Roraima and Tocantis. The population of the Amazon Legal increased from 7.3 to 13.2 million, GDP increased from \$2.2 billion to \$13.5 billion, and 33 million hectares of forest were converted to agricultural land between 1970 and 1985 (Andersen & Reis 1997). However land was often later abandoned and social indicator statistics showed that the Amazon still lagged behind the rest of Brazil in terms of income, education and life expectancy (Andersen & Reis 1997). The pattern of boom-and-bust in development is evident throughout the economic history of Brazil (Godfrey 1990, Macedo & Anderson 1993, Clough *et al.* 2009, Ahmed *et al.* 2013), with boom and bust cycles depending heavily on extractive industries (Godfrey 1990).

Development indices in the Amazon such as life expectancy, literacy and standard of living, often follow the economic cycles of boom and bust (Rodrigues *et al.* 2009).

2.5.3. Roads, logging and deforestation in the Amazon

It has been noted there is a strong link between development and deforestation, with economic development leading to increased demand for land and resources that in turn leads to a loss of forest and biodiversity (Wilkie *et al.* 2000, Southworth *et al.* 2011). Road and other infrastructure development is often perceived as the initial stage of development, opening access to remote areas for colonisation, agriculture development, resource extraction, and linked with these; deforestation (Chomitz & Gray 1996, Laurance *et al.* 2001, Perz *et al.* 2007, Laurance *et al.* 2009, Caldas *et al.* 2010).

Southworth *et al.* (2011) report that deforestation patterns often closely mirror the pattern of the road network and that deforestation rates drop with distance from main roads. However this drop of in deforestation with distance from roads is moderated by level of development, for example Acre, Brazil, has completed road paving and roads influence deforestation upto 45 km out from the road edge, whereas in Madre de Dios, Peru, where road paving is incomplete (therefore is considered less developed than Acre) deforestation extends 18 km out from the road edge (Southworth *et al.* 2011). This link between roads and deforestation highlights the important role that infrastructure development plays in determining the spatial patterns of deforestation and biodiversity loss.

The initial roads to develop in the Amazon are often logging roads, built to access timber. Timber is a huge global market with the top 100 forest and paper companies having a combined revenue of US$ 357 billion (0.05% of the global economy) in 2008 (Dauvergne and Lister 2012). Only 20% of Brazil's timber is bound for the global market, with much of the harvested timber being consumed locally, yet in 2006 it had a 3.9% share in global timber, pulp and paper exports (Sierra 2001, Dauvergne and Lister 2012) . Four out of the five largest merger and acquisition deals in the global forest market took place in South America, accounting for two thirds of the overall deal value in 2009 (Dauvergne and Lister 2012). Within Brazil, a single processing mill owned by Veracel (a venture of Fibria) produces 3,800 tonnes of pulp a day, supported by a 90,000 ha eucalyptus plantation. However, the number of plantations in South America is low, with the majority of timber being sourced from natural forests (Sierra 2001), in fact 90% of Brazil's timber is extracted from natural forests (Matricardi *et al.* 2005).The cost to harvest timber in Brazil is much lower than in many other places, for example in 2009 the Brazilian based company Fibria was able to produce wood pulp at US$ 222 per tonne, much lower than the global average of US$ 389 per tonne (Dauvergne and Lister 2012). Approximately 350 Amazonian tree species are commercially harvested, producing an estimated 24 to 28 million m^3 of roundwood timber annually, generating revenue of US$ 2.5 billion (Verissimo & Cochrane 2003, Merry & Amacher 2005, Arima *et al.* 2005). Timber extraction is also a source of employment in the region, in the state of Mato Grosso 38,000 people were employed in 2004 extracting and processing timber (Rodrigues-Filho *et al.* 2012).

Usually between 2 and 9 species are harvested per hectare of logged forest (Asner *et al.* 2006, Pereira *et al.* 2001, Broadbent *et al.* 2008). Although the majority of logging is selective, with only a few valuable trees being removed per hectare, the actual logging process can cause a great deal of damage to the surrounding forest, with upto 50% of the remaining canopy being damaged by logging operations (Matricardi *et al.* 2005). Until recently timber harvesting in Brazil has been a mosaic of small-holder deforestation, illegal logging on private and public lands and legal harvesting of private lands (Merry and Amacher 2005). Recently legal logging on public lands, through government ordained concessions has been allowed; with bidding on the first two concessions resulting in 144,800 ha of forest being concessioned by 2010, and a further 1,026,000 ha of forest becoming open to bidding in 2011 (SFB 2013). Each year over 1million hectares of standing forest are selectively logged, this is in addition to timber procured during deforestation activities (Verissimo & Cochrane 2003). 50% of timber arriving at sawmills within the Amazon is illegally harvested, generally from unclaimed public lands (Verissimo & Cochrane 2003), however upto 80% of logging activities may be illegal (Brack 2003). Even conservation areas with protected status are not safe from logging activity with 1,200 km^2 of logging observed in conservation areas between 1999 and 2002 (Asner *et al.* 2005). In addition to 'predatory' logging, essentially cut-and-run, other forms of illegal timber exist. For example, underreporting of harvest, undeclared or misreported harvest species, high-grading (where the best trees are removed, the remaining low value stock re-seeds the area, but the overall value is reduced because high-value species are no longer present or present in low proportions), tax and royalty evasion, subsequent environmental offences (e.g. not following a submitted management plan) and exceeding harvest boundaries. While citation of illegal logging is high, the probability of prosecution or conviction is low (Merry and Amacher 2005).

Logging operations in Brazil really began in the 1970s and timber was often close to mills with loggers usually travelling a few kilometres from mills to access timber. By the mid-1990s loggers regularly travelled over 100km to access desirable timber (Johns *et al.* 1996). This has inevitably led to extensive road network development by loggers to access timber stands and transport timber to mills. In addition, many new mills have been set up deeper into the forest frontier along the new networks (Merry & Amacher 2005). Johns *et al.* (1996) divide the logging process into five stages; 1) Bulldozers open a network of logging roads, 2) Patches of forest are cleared to serve as patios or log landings, 3) Trees are felled and bucked (limbs removed and logs are produced), 4) Logs are linked to a bulldozer or skidder, 5) Logs are skidded to landings before transport. Prior to this activity, roads are built from the existing road network out into forest areas of interest. Indeed in many frontier regions of Brazil it is logging activity that is the main cause of road construction. One study found two thirds of roads surveyed were built by loggers, often in exchange for logging rights on the land (Uhl *et al.* 1991).

Two main forms of logging exist, planned (also called reduced impact logging) and unplanned, which includes illegal logging. Planned logging involves additional effort in the form of, inventory mapping of trees, planning the most suitable locations for roads and landings, vine cutting in advance to reduce damage to surrounding trees (which may be harvestable later), planned directional felling, and planned extraction (Pereira *et al.* 2001). While this additional planning (compared to unplanned logging) costs approximately US$ 72 more per ha, this investment is offset by increases in productivity (15%), less machine time to build roads and landings (37%), increased timber hauled to landings (27%), and less timber wasted to poor felling techniques and unfound felled trees (26%) when compared to unplanned operations (Barreto *et al.* 1998). Further, planned logging operations generally have much lower environmental impacts than do unplanned operations, with smaller canopy openings, fewer non-harvest trees damaged, less ground disturbance from machinery and narrower skid trails (Johns *et al.* 1996, Pereira *et al.* 2001). Interestingly, fewer roads are built under planned (or reduced impact) logging operations compared to unplanned (and illegal) operations extracting the same volume of timber, for example Periera *et al.* (2001) found that roads covered 1.2% of the harvest area in an unplanned operation compared to just 0.6% in the planned harvest area, two years later the planned area had 1% road coverage and the unplanned area had 2%. The vast majority of logging, approximately 95%, is unplanned which causes more damage than planned logging (Johns *et al.* 1996, Pereira *et al.* 2001, Verissimo *et al.* 2002).

Even when forests are selectively logged with little environmental damage, many areas that are logged are often deforested within a few years (Asner *et al.* 2006), primarily because access is granted to agriculturalists, land prospectors, and colonists who utilise the roads built by loggers and cause deforestation and degradation (Fearnside 2007, Laurance *et al.* 2004). Logged regions are also at greater risk of further forest loss through fire risk caused by edges created by roads (Broadbent *et al.* 2008, Nepstad *et al.* 1999, Nepstad *et al.* 2001, Uriarte *et al.* 2012). This makes the road network a key factor in deforestation patterns, with studies showing that roads and deforestation are closely linked (Chomitz & Gray 1996, Laurance *et al.* 2001, Perz *et al.* 2007, Laurance *et al.* 2009, Caldas *et al.* 2010, Southworth *et al.* 2011). Consequently, roads have been found to be one of the most commonly used inputs in land use land cover (LULC) change models in the Amazon, with a recent review reporting 24 out of 35 published studies utilise information on roads as inputs to models

(Rosa *et al.* 2014), with roads determining the accessibility of land and the cost of transportation which in turn determines the viability of land use change. Most deforestation and LULC models treat road networks as a static pattern. However, it is widely acknowledged that road networks in the Amazon are highly dynamic with roads growing at rapid rates, with an average increase of 17,000 km of new road added per year (Brandão and Souza 2006, Ahmed *et al.* 2013). There are many difficulties associated with predicting this largely anthropogenic phenomenon that is subject to many idiosyncratic events, which is possibly why LULC models treat it as static. Thus predicting road network development remains a challenge in predicting future deforestation (Barlow *et al.* 2011).

2.6. Conclusion

The Amazon rainforest is highly diverse, productive and offers vast array of ecosystem goods and services. It is subject to many pressures including, extractive industry, resource exploitation, poor governance, a changing climate and infrastructure development. Given the importance of this ecosystem, the fact that it has a rapidly growing road network (that is largely unplanned), and the extensive negative ecological effects a road network can cause, it is imperative that we understand and are able to predict the spatio-temporal dynamics of road network growth in this globally important system. To date four different road models have been used to predict the growth of the Brazilian road network (Arima *et al.* 2008, Soares-filho *et al.* 2004, Jiang 2007, Walker *et al.* 2013) all of which utilise least-cost path algorithms to determine the path of developing roads, however only two of these have been validated (Arima *et al.* 2008, Walker *et al.* 2013), neither of which are available for general use (i.e one cannot use the models to generate predictions of road development for subsequent use, in land use models for example). In order to further the field of infrastructure development modelling, in this thesis, I utilise two distinct approaches to generate amazon-wide road models that could potentially be used to estimate the future of road development. I aim to subject these models to rigorous validation to establish models that are useful but also transparent in terms of predictive capabilities. In addition to presenting the models I will also present supporting work that contribute to the models' development. It is hoped that the results of this thesis can be used to further improve our understanding and ability to predict road development and its ecological implications in this valuable ecosystem.

Chapter 3: **Can roadless volume predict patterns of biodiversity? A test using birds in the central Amazon**

3.1. Abstract

Roads have significant negative impacts on biodiversity; their presence in landscapes leads to edge effects and habitat loss, fragmentation and environmental perturbation. However, little work has been done to evaluate the direct links between road networks and biodiversity loss, as opposed to the indirect effects of road networks via habitat loss. We compared forest bird species richness in the municipalities of Santarém and Belterra in Pará state, Eastern Brazilian Amazon, with a road network metric called 'roadless volume' at multiple spatial scales ranging from local landscapes (28 ha) through to natural water catchments (averaging 3721 ha). We found a significant positive relationship between roadless volume and forest bird richness at all three spatial scales under investigation, indicating a negative relationship between species richness and road occurrence. Roadless volume was also positively correlated with the average number of unique species recorded within each site. Regression between roadless volume and DCA scores (De-trended correspondence analysis) showed that forest bird community composition is significantly negatively related to roadless volume and sites with similar roadless volumes had similar community compositions. Therefore roadless volume not only influences species richness but also community composition. We found no significant correlation between roadless volume and forest cover, thus we suggest that road networks impact biodiversity independently of habitat cover. Possibly because roadless volume is a proxy for disturbance level and fragmentation, which may not be captured by total percentage habitat cover. Thus roadless volume is able to disentangle the impacts of road networks and habit cover, which may independently influence Amazonian biota.

3.2. Introduction

Infrastructure developments, particularly roads, are a ubiquitous feature of landscapes modified by humans. While roads and roadsides already account for 1-2% of the land surface in many developed countries (Forman 1998), the rates of road expansion are fastest in the developing tropics and emerging economies where they are given high priority by governments to encourage growth and reduce poverty through increasing spatial connectivity, aiding travel, helping establish land claims and facilitating the extraction of resources (Munnell 1992, Calderon & Serven 2004, Straub 2008, Perz *et al.* 2012).

Despite the irrefutable socio-economic benefits that roads bring to humans, they often result in negative impacts on native biota (Forman & Alexander 1998, Spellerberg 1998, Fahrig & Rytwinski 2009, Laurance *et al.* 2009, Perz *et al.* 2012). Many road impacts cause changes to biodiversity richness and composition (Wilkie *et al.* 2000, Forman *et al.* 2003, Spooner & Smallbone 2009). Such road impacts can include roadkill, loss of habitat, and the formation of barriers to animal dispersal and gene flow. Road crossing avoidance has been observed in many groups including invertebrates (Keller & Largiader 2003, Bhattacharya *et al.* 2003), amphibians (Marsh *et al.* 2005), reptiles (Shepard *et al.* 2008), birds (Lees & Peres 2008, Laurance *et al.* 2004, Tremblay & St Clair 2009), and mammals (Richardson *et al.* 1997, Dyer *et al.* 2002, Rico *et al.* 2007, McGregor *et al.* 2008). Road avoidance behaviour affects species' distributions causing range restrictions by fragmenting populations of non-vagile organisms. Roads also affect biodiversity through reducing habitat quality, facilitating human access to frontier areas, fragmenting habitats and creating edge effects at road-habitat boundaries (Forman & Alexander 1998, Keller & Largiader 2003, Shyama Prasad Rao & Saptha Girish 2007, Jaeger *et al.* 2005).

Birds are particularly sensitive to roads, with studies showing roads negatively affect their reproductive success (King & DeGraaf 2002), occurrence (Kuitunen *et al.* 1999, Ortega & Capen 1999, Clark & Karr 1979, van der Zande *et al.* 1980, Reijnen & Foppen 1991, Forman & Deblinger 2000), movement (Develey & Stouffer 2001, Laurance *et al.* 2004, Lees & Peres 2008) and even vocal activity (Brumm 2004, Fuller *et al.* 2007, Slabbekoon & Ripmeester 2008). Avian dispersal capacity, and hence gap-crossing propensity, is highly species/guild specific; while some species routinely undertake migrations between hemispheres, other volant species of similar body size are physiologically incapable of flying more than 100m (Moore *et al.* 2008, Bairlein *et al.* 2012).

Quantifying the link between roads and bird diversity is especially important in tropical deforestation frontiers, where high biodiversity is particularly vulnerable to environmental perturbations caused by human activities (O'Neill 1976, Wright & Muller-Landau 2006, Gardner *et al.* 2009). Here, we focus on the Brazilian Amazon, which is the world's largest remaining area of tropical forest (Foley *et al.* 2007). This region contains over 1,300 bird species (Marini & Garcia 2005). Road networks are expanding at rapid rates in the Brazilian Amazon, growing by almost 17 000 km per year between 2004 and 2007 (Ahmed *et al.* 2013). Following this expansion there has been extensive concomitant habitat loss and fragmentation. Key to this growth in infrastructure are government led development plans, including 'Operation Amazonia', 'National Development Plan' in the 1960/70s, the more recent 'Brasilem Acao', the 'inter-ocean highway' and 'Avanca Brazil'. All of these plans heavily feature infrastructure development (Andersen & Reis 1997, Carvalho *et al.* 2002, Killeen 2005) and have led to the construction of major highways such as the BR-10, BR-163, BR-219 and BR-319 that have largely defined development patterns in the region.

Individual states in the Brazilian Amazon have their own transport plans, such as PELT-Pará for the state of Pará, in which our study site is located. Based on predictions of economic growth and transport logistics, it is estimated that over 27,000 km of roads will be constructed, improved (paved), expanded (more lanes), extended and maintained in Pará by the year 2031 (PELT-Pará 2012). This extensive development plan exemplifies the importance placed on road construction to maintain economic growth and highlights the degree of expected development within the region. It also raises concerns over the lack of importance placed on the biodiversity impacts of these infrastructural changes, given there is no indication of any environmental assessment within the PELT-Pará plan. Given the expected development of roads with in the Amazon region it is of utmost importance that the link between roads and biodiversity is quantified to facilitate accurate assessments of potential impacts.

To date, attempts to quantify the relationship between roads and biodiversity have relied on two alternative analyses. The first uses the direct relationships between spatial patterns of road networks and biodiversity, which allows us to make useful predictions about future biodiversity changes based on expectations of road network expansion. A key example of this is the GLOBIO model, developed by the United Nations Environment Program (UNEP), which relates the biodiversity metric Mean Species Abundance (MSA) to distance from roads (UNEP GLOBIO 2001). The second way to assess the relationship between roads and biodiversity is indirect, with patterns of road development being related to habitat loss which is then used to estimate species loss. Examples of this approach include the Millennium Ecosystem Assessment which bases predictions of biodiversity change by combining the outputs of IMAGE2.2, a model that predicts changes in land use, with the species-area relationship (MacArthur & Wilson 1967).

Roadless volume (Watts *et al.* 2007), is a relatively new metric of road network density that accounts for the exact spatial pattern of roads within an area of interest. It is the amount of space there is between roads, with the value of that space weighted by distance to the nearest road, such that areas that are less disturbed by roads have a higher roadless volume value. Roadless volume is simple to calculate in a Geographic Information System, making it an attractive option for estimating the extent to which roads pervade landscapes at multiple spatial scales. Here, we use data on bird species richness collected at multiple spatial scales in a deforestation frontier region in the state of Pará to determine the relationship between roadless volume and biodiversity. Our goals are to quantify the relative importance of roads *versus* habitat amount in determining biodiversity patterns in this tropical region, with a view to using roadless volume to estimate the impacts of historical road network expansion on biodiversity in a region where road coverage is expanding rapidly.

3.3. Methods

Our study is based in the municipalities of Santarém and Belterra, Pará state, Brazil (Figure 3.1). The study site was divided into 286 natural water catchments of average size 3721 ha (SD = 2747), delineated using a DEM (digital elevation model, based on NASA SRTM data) and SWAT (Soil and Water Assessment Tool) in ArcGIS 9.3. All analyses detailed were carried out in the statistics program R 2.10.1 (R Development Core Team 2009) unless otherwise stated.

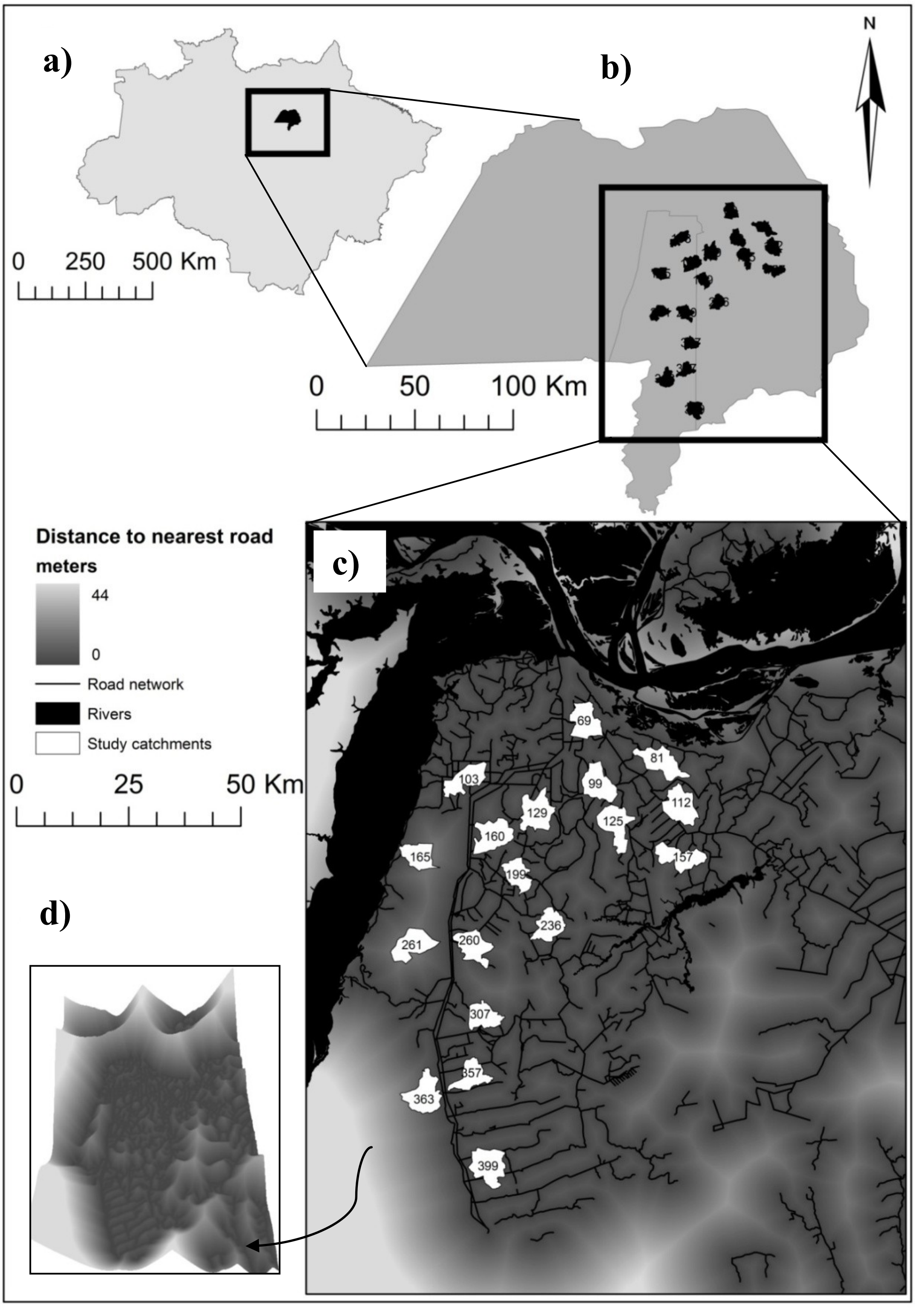


Figure 3.1. Study site a) Location within the Brazilian Amazon Legal, b) Municipios, Santerem and belterra, with 18 study catchments highlighted in black, c) Distance to nearest road, surface calculated over 30m grid, with river location and roads in 2008 shown in black. 18 study catchments are highlighted in white (numbered as in Lees *et al.* 2013), d) Roadless volume calculated over an equal area 60m grid.

3.3.1. Bird sampling

A subset of 18 catchments (Figure 3.1) were chosen to represent a gradient of forest cover from 11% to 100%. A stratified-random sampling design was employed to help ensure a representative assessment of bird species richness in each land cover within the catchments, resulting in between 6 and 12 transects per catchment. Sample transects were distributed randomly across the catchments to increase the likelihood that important internal heterogeneities in land cover were captured. A minimum separation distance rule of 1500 m between transects was employed to maximize independence between sampling points (see Gardner *et al.* 2013, Lees *et al.* 2013 for more details). Bird richness surveys were conducted by Alex C. Lees, Nargila G. Moura, Christian B. Andretti, Bradley J.W. Davis & Edson V. Lopes between 16 October 2010 and 8 February 2011. Two repetitions of three fixed width (75m) 15-minute point counts per transect were conducted, points were 150m apart (see Figure 3.1 for location of study catchments. All birds observed (based on sightings or call recognition) were recorded, however for this study only forest bird observations were used. Birds were classified as forest species based on the classifications of Henriques *et al.* (2003) and personal observations of those conducting surveys if the species occurred in extensive areas of intact forest. In order to evaluate sample representation, estimates of bird species richness (Chao 1) were calculated using EstimateS software (Colwell 2009) individually for each catchment and for total samples at catchment scale.

3.3.2. Calculating roadless volume

Two satellite images from Landsat location 227/062, covering a period between 2000 and 2008, were manually digitised to generate road network maps following the methods described by Brandão & Souza (2006) and used to calculate roadless volume in ArcGIS 9.3.

As the maps were digitised from satellite imagery we were unable to differentiate between road types (e.g. paved *versus* unpaved), thus all visible roads were included in the maps and treated equally. For catchment scale analyses, we first generated a ‘distance to nearest road’ raster grid, using the ArcGIS Euclidean Distance tool at 30m resolution (Figure 3.1c), then using the Hawth’s tools Zonal Statistics tool (Beyer 2004), we calculated the sum of raster cells for each catchment, i.e. the sum distances to nearest road. These values were divided by catchment area to give a standardised metric of roadless volume for each catchment. The catchment forms a natural ‘footprint’ to use as an area upon which to base roadless volume calculations. At point and transect scale analyses, buffers were applied to produce a ‘footprint’. For point scale analyses we used a 75m buffer to avoid overlap with adjacent points (150m separation) and to capture the width of the counts. We used multiple buffers (100-10,000m) for transects to determine the ‘best’ scale at which to calculate roadless volume at the transect scale. The slope and r^2 from linear regressions of roadless volume vs. bird richness were plotted against buffer size (following the protocols presented by Steffan-Dewenter 2002) to select the buffer size that had the highest explanatory power, which was found to be 2000m. Roadless volume was log transformed at all three scales (catchment, transect, point).

A strong relationship between roads and habitat loss (deforestation) may confound any relationship between roads and species richness, thus we investigated the degree to which roadless volume and percentage forest cover were correlated at the catchment scale. Percentage forest cover was log transformed for the Pearson’s correlation. The relative importance of roadless volume and habitat cover in relation to species richness was determined using variance partitioning on linear models with roadless volume and percent forest cover as explanatory variables.

3.3.3. Comparing roadless volume with bird richness and composition

Linear regression was used to determine the relationship between roadless volume and forest species richness at all three sampling scales (point, transect & catchment). Because samples were nested, we used mixed effects models to take the hierarchal nature of the data into account. Linear mixed effects models (LMEM) were generated in R using the 'nlme' package (Pinheiro *et al.* 2012). We modelled point scale species richness against roadless volume as a fixed effect and scale as a random effect with two levels of nesting (point within transect within catchment). For transect level richness an LMEM with only one nested level (transect within catchment) could be used, with transect level roadless volume as a fixed effect and scale as a random effect.

To examine how roadless volume influences community composition at the catchment scale, the number of log-transformed unique species (i.e. counted once per catchment; used here as a proxy for rare species) for each catchment was regressed against roadless volume. We also quantified communities using de-trended correspondence analysis (DCA) to represent the community composition pattern of bird communities within catchments. The first two DCA axes explained 28% and 12 % of the variance in community composition among catchments respectively, while the third and fourth axis each explained 0.8%. Consequently, we regressed the DCA axis scores from the first two axes alone against roadless volume in order to investigate if overall forest bird community changes with changing roadless volume.

We extrapolated the relationship between roadless volume and forest species richness at the catchment scale to predict the species richness of forest birds in each of the 286 catchments delineated using DEM and SWAT. We used the same ‘distance to nearest road’ surface calculated over a 30m grid for initial analyses to calculate the roadless volume for all 286 catchments in 2008. A second ‘distance to nearest road’ surface was generated for the earliest digitised road map we had available (2000) and the roadless volume for each catchment in 2000 was calculated. Based on the relationship established between roadless volume and species richness at the catchment scale we estimated species richness in each catchment both for 2000 and 2008, and compared the two to estimate the number of local extinctions over the eight year period following the road network expansion across the study site.

3.4. Results

A total of 11,028 individual birds from 384 species were recorded during the timed point counts, of these 8,743 detections from 298 forest bird species were selected for the analyses. Chao1 estimates of forest bird species richness were generally slightly higher than observed forest bird richness, with an estimated species richness of 331 (CI=312-373) at the catchment scale. Richness estimates suggested sampling had captured an average of 79 % of species within each catchment (SD = 5.7).

Roadless volume was positively, but not significantly, correlated with percentage forest cover at the catchment scale ($r=0.447$, $df=16$, $p=0.063$).The regression model incorporating both variables (roadless volume and percent forest cover) performed well, explaining 91 %

of the variation in bird species richness. There was no significant interaction between roadless volume and percent forest cover (F=53.49, p=0.84, df=14). roadless volume alone explained 15% more variance in species richness compared to percent forest cover alone, however these two models were not significantly different from each other, and both performed significantly worse than the model with both variables included based on anovas (Table 3.1).

Table 3.1. Regression models and variance partitioning of bird species diversity regressed against roadless volume (RV) and percentage forest cover (%FC). All models shown are significant with p<0.001. Shown in brackets is the individual contribution to variance explained of RV and %FC, model degrees of freedom (df), whether the models are significantly different from each other and the full model (RV+%FC), which variables are significant and model AIC.

model predictors	r^2	d.f.	Sig.diff compared to RV+%FC?	Sig.diff compared to RV?	Sig variables	AIC
RV+%FC	0.91	15	N/A	Yes	RV & %FC	133
RV	0.72 (0.34)	16	Yes	NA	RV	152
%FC	0.57 (0.19)	16	Yes	No	%FC	160

Linear regression models showed a significant positive relationship between roadless volume and species richness at point (r^2= 0.12, slope=1.90, df=521, p<0.001), transect (r^2=0.25, slope= 12.9, df=169, p<0.001), and catchment scale (r^2=0.73, slope=324, df=16, p<0.001). There is a clear scale effect with the relationship between roadless volume and species richness strengthening with increasing spatial scale (Figure 3.2). These results held when tested with mixed effects models, indicating the trend is robust to analysis method (Table 3.2).

Roadless volume also exerted a significant effect on species composition, reflected in analyses of the overall community composition (DCA ordination) and the number of unique species per catchment. There was a significant negative relationship between roadless volume and DCA axis1 scores (r^2= 0.30, slope = -8.4, df=16, p<0.05, Figure 3.3a), There was a positive relationship between roadless volume and DCA axis2 scores, however this relationship was not significant (r^2= 16, slope= 4.6, df=16, p=0.09). We also found that the number of unique species present in any given catchment increased with increasing roadless volume (r^2=0.32, slope= 5.8, df=16, p< 0.05, Figure 3.3b).

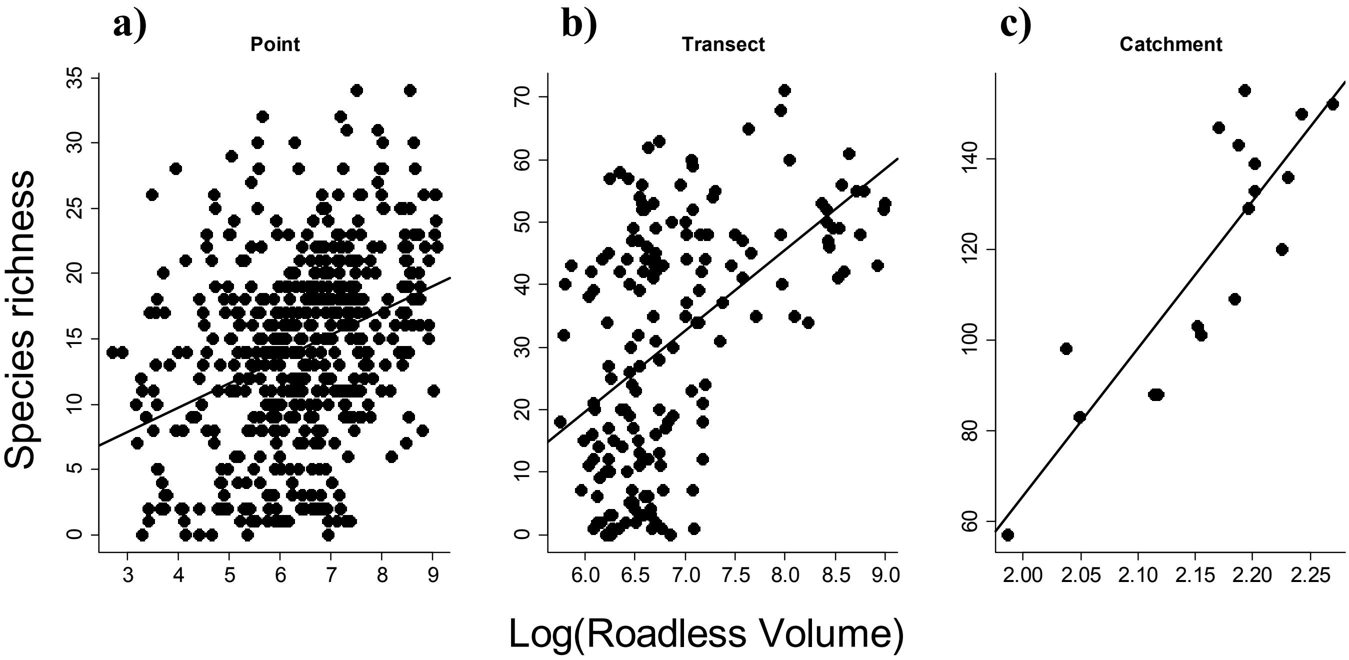


Figure 3.2. Roadless volume and forest bird species richness across multiple scales; a) Point scale, b) Transect scale, and c) Catchment scale.

Table 3.2. Mixed effects models results, with species richness as a function of roadless volume (RV) at two scales (point and transect).

	Variable	Point | transect | catchment	Transect | catchment
Fixed Effects (RV)	Slope	1.60	11.98
	SE Slope	0.26	2.03
	Intercept	4.08	-15.97
	d.f.	503	151
	t-value	6.08	5.89
Random Effects (location)	Intercept	2.63	15.98
	Slope	2.63	1.99
	Residual	6.33	14.8
	p-value	<0.001	<0.001
	Number of observations	523	171
	Number of groups	19	19

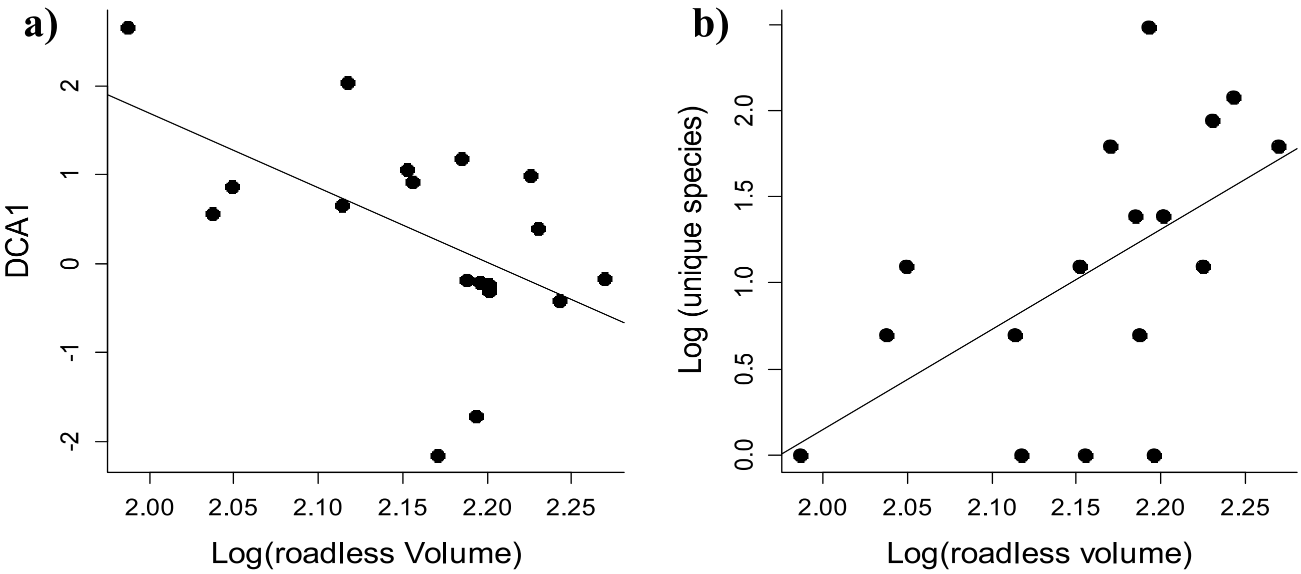


Figure 3.3. Roadless volume and community composition. a) The relationship between roadless volume and DCA axis1 scores representing bird community composition at the catchment scale. b) The number of unique species present at any given watershed increases with increasing roadless volume.

Areas of predicted high bird species richness and high roadless volume corresponded to the remaining areas of intact forest – a forest reserve (FLONA TAPAJOS) along the western border and areas of unprotected primary forest in the south of the study region. These areas retained high avian species richness throughout the 8 year period (Figure 3.4, a, b). Between 2000 and 2008 there were 2773 km of new roads, mostly in the middle of the study region along the east bank of the river around the existing year 2000 road network (Figure 3.1, Figure 3.4d). The model predicted an average bird species richness of 141 (SD=24) per watershed in 2000, and that this had reduced by an average of 18 (SD=19) species per catchment in the year 2008. Spatial pattern of local extinction followed that of road expansion, and resulted in particularly heavy losses of over 50 species in 24 catchments (highlighted in Figure 3.4), with a maximum loss of 96 species in one location. Catchments with particularly heavy losses were concentrated along the river. No roads were detected in the southern most catchments in either year; unsurprisingly no species were lost in these areas. Only 31 of the 286 catchments in the study region lost no species over the eight year time frame.

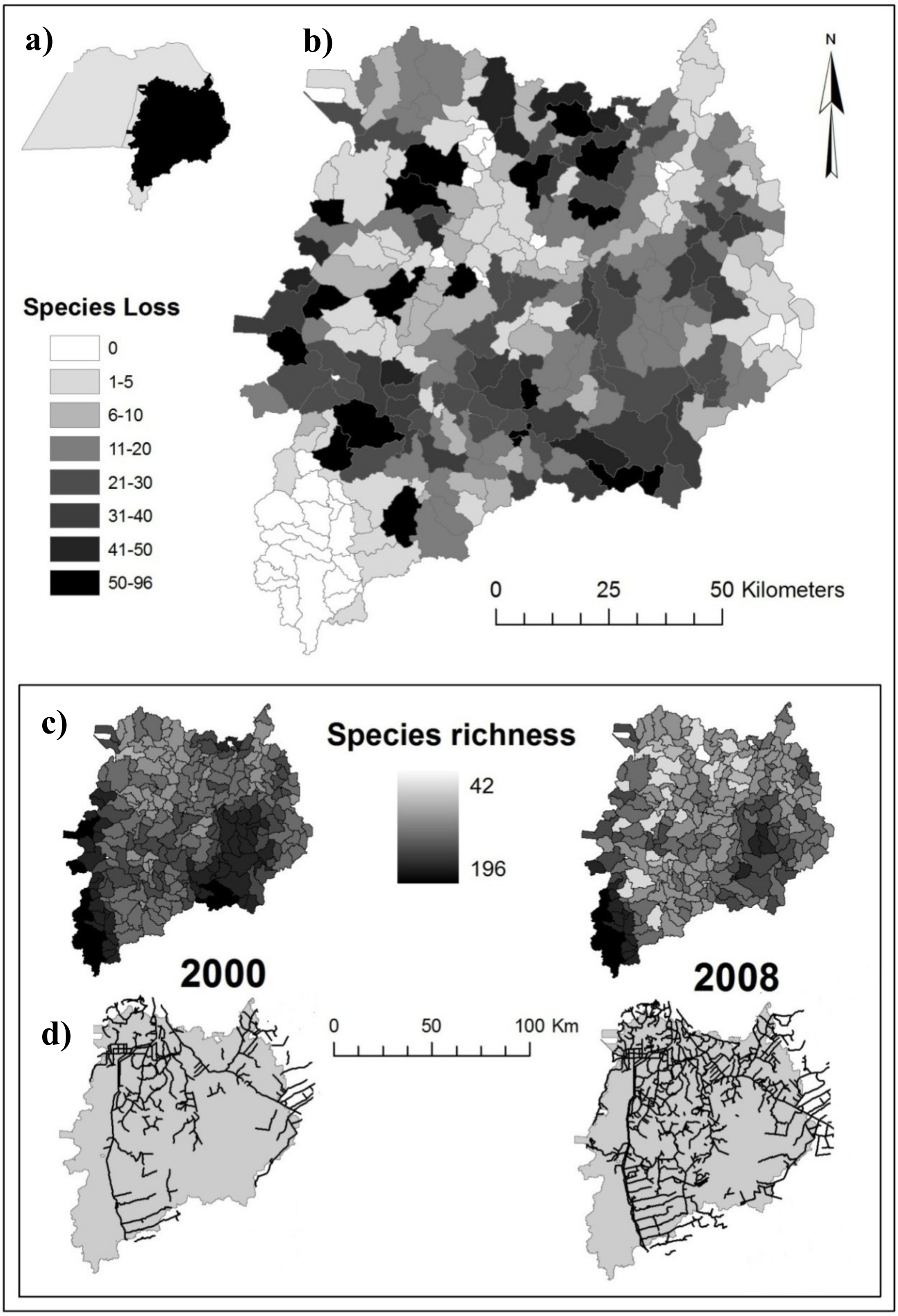


Figure 3.4. Estimated species loss between 2000 and 2008 across 286 water catchments in Santarém and Belterra based on changes in roadless volume. a) Catchment area location within Santarem and Belterra, b) Estimated bird species loss, c) Estimated species richness in 2000 and 2008 based on roadless volume from corresponding years. d) Road network in 2000 and 2008

3.5. Discussion

We found there is a clear positive relationship between roadless volume and bird species richness and composition, which increased in strength with larger sampling scales (Figure 3.2), which have greater levels of species representation. Based on the relationships we detected, we estimate water catchments in the state of Pará have lost an average of 18 species between 2000 and 2008, with just 11% catchments experiencing no loss of species.

By disentangling the impacts of roads from the impacts of habitat loss, using a road metric that was not significantly correlated to habitat cover, we were able to explore the relationship between the road network and bird biodiversity and conclude that road networks impact biodiversity independently of habitat loss. Road networks are likely to affect bird species richness and community composition in two main ways. Firstly by fragmenting the habitat and restricting movement between patches, even very narrow roads can disrupt the movement of many forest-associated species, particularly terrestrial insectivorous passerines (Lees & Peres 2008). Secondly environmental changes, such as new habitat edges and subsequent edge effects, alter which species select or are able to remain in a given area. These effects can occur with almost insignificant amounts of habitat loss, ensuring that the mechanisms by which roads impact biodiversity do not necessarily rely on the outright loss of habitat. Moreover, the presence of roads may be a good proxy for more cryptic forms of disturbance, such as forest degradation from logging, fire and hunting (Peres *et al.* 2006). The relationships we found between roadless volume and biodiversity may, in part, reflect this correlation between forest disturbance and road density that again is not necessarily correlated with outright habitat loss.

As roads are built, it is likely that sensitive specialist forest species are lost from the community first, followed by the loss of increasingly generalist species accompanied by an increase in the number of edge and gap species. Laurance (2004) and Laurance *et al.* (2004) found that edge and gap specialist bird species were more prevalent close to road edges than deeper inside the forest, whereas the abundance of specialist insectivorous, terrestrial and solitary species declined close to roads. The number of unique species present in a given catchment increases with increasing roadless volume, possibly because species with very low abundances are highly sensitive to road presence and so are unlikely to be detected in catchments with high road densities. While the degree of variance in community composition and unique species explained by roadless volume is fairly low (30% and 32 % respectively), it is promising that a trend is detected. Suggesting that roadless volume can indeed detect trends that have been observed in the field, it would be an interesting area of future work to compare this with other road metrics and determine which are the best predictors. Further, investigations into community composition would be beneficial as the currect scores of 28% and 12% are rather low (which is why the second axis (12%) was not used in regression analyses).

We found that roadless volume strongly influences both species richness and community composition at catchment scale, giving us confidence in using change in roadless volume as a suitable predictor of catchment-scale changes in forest bird species richness. Given that the Brazilian road network is likely to continue to develop in the foreseeable future (Ahmed *et al.* 2013), with increased investment in development plans from local (eg PELT- Pará) to continental scales (e.g. IIRSA; Initiative for the Integration of the Regional Infrastructure of South America), it is imperative that we are able to make forecasts of potential biodiversity changes as a result of road development. The strong relationship between bird species richness and roadless volume provides one avenue for making such forecasts, making it possible to examine the potential impacts of competing road layouts on biodiversity to help best design road networks of the future. Catchment scale analyses represent an appropriate spatial scale for these analyses given the effects of local biodiversity on the provision of ecosystem function and services at local scales (Wearn *et al.* 2012).

Extensions to the road network of 2,773 km in Pará resulted in projected local biodiversity losses. These projected losses are likely to increase further under local development plans which include a 27,000 km road network upgrade by 2031. This region is a microcosm of change in the Amazon. Given the expected expansion of roads across the Amazon (Kileen 2005, Ahmed *et al.* 2013) we can assume these local losses are likely to reflect losses on a wider scale. We have presented a method for quantifying the biodiversity impacts of infrastructure development plans that is a relatively simple exercise once spatially explicit biodiversity data has been obtained and provided appropriate roadmaps are available. Furthermore, predictions of the nature of future road networks based on development plans such as PELT- Pará, or modelling, can be used to assess changes in species richness that may be of use to conservation planners or policy makers. However, caution should be excersised as the methods and projections presented here have not been validated, and validations against real loss data both in the study region and else where would be necessary to determine the accuracy and trahnsfereability of this approach.

Chapter 4: **Spatial pattern of standing timber value across the Brazilian Amazon**

4.1. Abstract

The Amazon is a globally important system, providing a host of ecosystem services from climate regulation to food sources. It is also home to a quarter of all global diversity. Large swathes of forest are removed each year, and many models have attempted to predict the spatial patterns of this forest loss. The spatial patterns of deforestation are determined largely by the patterns of roads that open access to frontier areas and expansion of the road network in the Amazon is largely determined by profit seeking logging activities. Here we present predictions for the spatial distribution of standing value of timber across the Amazon. We show that the patterns of timber value reflect large-scale ecological gradients, determining the spatial distribution of functional traits of trees which are, in turn, correlated with timber values. We expect that understanding the spatial patterns of timber value across the Amazon will aid predictions of logging movements and thus predictions of potential future road developments. These predictions in turn will be of great use in estimating the spatial patterns of deforestation in this globally important biome.

4.2. Introduction

The Amazon is the largest remaining area of tropical forest (Foley *et al.* 2007), containing half of the world's tropical forest biome (Betts *et al.* 2008), covering an area of nearly 5 million km^2 (Moran 1993) and accounting for approximately 10 % of the Earth's terrestrial net primary productivity and biomass (Melillo *et al.* 1996, Malhi & Grace 2000). It is a highly biodiverse system (Dirzo & Raven 2003) housing a quarter of all global biodiversity (Betts *et al.* 2008). However it is also a system under threat, with an average of 19,500 km ^{2} of forest cleared each year between 1996 and 2005 (Nepstad *et al.* 2009). The spatial patterns of deforestation are determined largely by the patterns of roads that open access to frontier areas, leaving them susceptible to colonisation and further development (Fearnside 1987, Laurance *et al.* 2000, Verissimo *et al.* 1995, Mertens *et al.* 2002, Perz et al 2007, Laurance *et al.* 2009, Caldas *et al.* 2010). Kirby *et al.* (2006) showed that distance from roads is in fact the strongest predictor of deforestation in the Amazon, and Southworth *et al.* (2011) reported that deforestation patterns often closely mirror the pattern of the road network with deforestation rates falling with distance from main roads.

The expansion of the secondary road network in the Amazon is primarily driven by the logging sector (Arima *et al.* 2005) and logging is a huge industry in the Amazon, with an estimated US$ 2.5 billion of timber extracted each year (Arima *et al.* 2005). Estimates suggest that the amount of forest that is clear cut each year is matched with an equal area being selectively logged each year; approximately 10,000-15,000 km^2/year (Nepstad *et al.* 1999, Laurance *et al.* 2002). Even selectively logged forests can lose more than 40 % canopy cover through damage to surrounding trees during extraction and increased fire risks (Nepstad *et al.* 1999). Additionally, roads used by loggers to access timber also serve to open up frontier areas to colonists who further degrade and deforest (Caldas *et al.* 2010, Laurance *et al.* 2009, Perz *et al.* 2007).

A key goal of the logging sector, indeed any economically driven sector, is to maximise profits. This forms the basis of many land cover/land use change models, which assume a desire to maximise profit and use profit maximisation to determine potential land uses (e.g. Evans *et al.* 2001). There are two key aspects to profit; revenue and costs. In the logging industry the amount and value of timber extracted determines revenue. To accurately model deforestation driven by logging, it is, then, important to know the spatial distribution of timber values. The location of valuable timber is important because extraction is usually selective (Verissimo *et al.* 2002, Asner *et al.* 2004a) and loggers ideally want to harvest areas that yield the highest profits, i.e. are situated on high density, high value timber and that are accessible with the least cost.

Knowing the economic value of forests is important for conservation as well. For example, Verissimo *et al.* (2002) suggested locations for sustainable 'Flona' (national forests that allow sustainable logging) by combining data on protected areas, human occupation and forest value. They identified locations that would be economically viable to harvest but that would also provide biodiversity protection. The suggested a set of locations that covered 34 % of Amazonian forest, of which 38 % was also of high conservation priority. Further, the Amazon offers a host of ecosystem services, from climate regulation, water regulation and carbon storage, to forest products (Foley *et al.* 2007, Fearnside 2005, Bradshaw *et al.* 2007, Asner *et al.* 2004b, Bonan 2008, Malhi *et al.* 2008, Montagnini & Jordan 2005). In order to calculate a true cost-benefit analysis of logging in these forests, the value of the ecosystem services provided by the standing forest needs to be quantified and compared with the timber values obtained by felling.

There are various estimates of the timber value in the Amazon; some specific e.g. $15.4 billion (Merry *et al.* 2009), and some less specific e.g. 'several trillion dollars' (Uhl *et al.* 1997). However, timber value across the Amazon is difficult to predict because extensive surveys are labour and cost intensive, thus timber value estimates are often based on modelling. Often when looking at forest value, timber values are estimated in terms of net value (profit). For example, Stone (1998) modelled the net value of timber using three price classes (valuable ~>300, medium ~ 200-300, and low value ~100-175 US$/$m^3$) as a decaying function of greater distances from sawmills reducing net value, with the assumption that loggers will extract valuable timber from further away. This approach was also used by Verissimo *et al.* (2002) and Merry *et al.* (2009) who built on the work of Stone (1998), making their spatial models more detailed in terms of industry behaviour.

Although spatial profitability has been modelled across the Amazon, it is interesting to know how much the forest is worth in terms of standing timber value. Therefore we have modelled timber value across the Amazon such that valuable tree stands are shown as valuable irrespective of extraction costs. We used ordinary kriging to generate a value map from RADAMBRAZIL survey data for 11 tree genera that are economically important. Kriging methods have been used before in the Amazon region to estimate the spatial distribution of tree diversity (ter Steege *et al.* 2003), tree species distribution (Prates-Clark *et al.* 2008) and timber density (Arima *et al.* 2008). Our approach extends that of Arima *et al.* (2008) who used kriging to estimate total timber density as a way of determining locations for logging road destinations. Here, we combine data on timber density and value to generate a map of potential timber revenue for the Amazon.

4.3. Analyses

The RADAMBRAZIL (IBGE 2011) survey is a selection of surveys carried out between 1968 and 1978 which aimed to map the natural resources of Brazil. Among the data collected on soils, geology and potential land uses, an extensive survey of vegetation was also carried out. The RADAMBRAZIL data set contains information on a total of 89 families and 513 genera of trees, recording the timber volume of individual trees within plots of known location. We aggregated species data by genus for 2465 RADAMBRAZIL forest plots across the Brazilian Amazon. Timber properties, and thus value, are inevitably heterogeneous within and between groupings; genera within families, species within genera and individuals within species will show variation. For example, within the *Tabebuia* genus, which is generally known for desirable hardwood timber such as *T. guayacan*, one may also find medium weight wood species such as *T. roseo-alba* or even light weight wood species such as *T. cassinoides* (Gentry 1992). However, it was felt that genus was an appropriate level at which to carry out analyses relating to timber values because 71 % (Baker *et al.* 2004) to 74 % (Chave *et al.* 2006) of variation in wood density measures among species is explained by genus affiliation, whereas only 25 % to 34 % is explained by family affiliation.

A list of the top 15 Amazonian tree genera harvested in terms of volume exported (in $1000m^3$) and the average price per cubic metre, in US$, was obtained from ITTO (International Tropical Timber Organisation, 2011). The average value for each genus (US$/$m^3$) was obtained by averaging the total export value (US$/$m^3$) over the two years for which data were available (2006 and 2007). We were able to obtain data from both RADAMBRAZIL and ITTO for a total of 11 genera (*Dicorynia, Goupia, Hymenaea, Hymenolobium, Manilkara, Mora, Nectandra, Peltogyne, Swartzia, Swietenia, Tabebuia*), which were used in all subsequent analyses. We multiplied the average export value by the total volume of timber recorded at each RADAMBRAZIL location to obtain a timber value for that genus in that plot. RADAMBRAZIL forest plots were 1 ha in area, so our resulting timber values are in units of US$.$ha^{-1}$.Genus value estimates were summed across all 11 genera in each plot to obtain a total timber value (US$.$ha^{-1}$). Total value data was log-transformed to make it normally distributed (all analyses were done with log-transformed total value data, unless otherwise stated). Timber value data from ITTO was for processed sawn wood, whereas the RADAMBRAZIL timber volume data was for standing timber. Timber volume is lost at many stages of the logging process, meaning that 1 m^3 of timber represents a different 'volume' in the ITTO and RADAMBRAZIL datasets. Measurements of timber volume loss in sawmill operations based in the Eastern Amazon state ofParásuggest a conversion efficiency of 35 % from raw timber to sawn wood (Gerwing *et al.* 1996). To account for this loss of timber volume and to standardise 'volume' measurements among the two datasets, we multiplied all values derived from RADAMBRAZIL data by a scaling factor of 0.35.

Ordinary kriging was carried out in ArcGIS 9.3 using the Geostatistical package suite of tools. We initially modelled each of the 11 genera separately (Figure 4.1), before combining values from all genera within plots (Figure 4.2) to interpolate total value across the Brazilian Amazon. Output values were anti-logged to give true US$/ha values of timber, after which a forest mask (based on Prodes satellite data from INPE) was overlaid on the output map to remove extraneous timber values from areas known to be non-forest and/or previously deforested. We also calculated and present the spatial pattern of error in the kriged values. Because our model was conducted on log-transformed data, we present log-transformed timber values and the standard error associated with that value (Figure 4.2 c,d). An assumption of kriging analysis is that values in locations that are closer in space are more similar than those that are further away. To test for this spatial autocorrelation, we calculated Moran's I in R 2.10.1 (R Development Core Team, 2009) using the 'pgirmess' library (Giraudoux 2011) for total timber value across the Amazon. Moran's I statistics showed that locations up to ~200 km were strongly positively correlated (Moran's I=0.26-0.19, $P<0.001$), after which the correlation coefficient drops to < 0.04 and is generally non-significant. Furthermore, we validated the un-scaled kriging predictions by using linear regression to model predicted against observed timber value. There was a significant, positive relationship ($r^2=0.31$, $p<0.001$, df=2463) and the regression line did not differ significantly from a 1:1 relationship (slope = 1.02, 95% confidence interval 0.95 – 1.09), indicating that the predictions of timber value by kriging were robust.

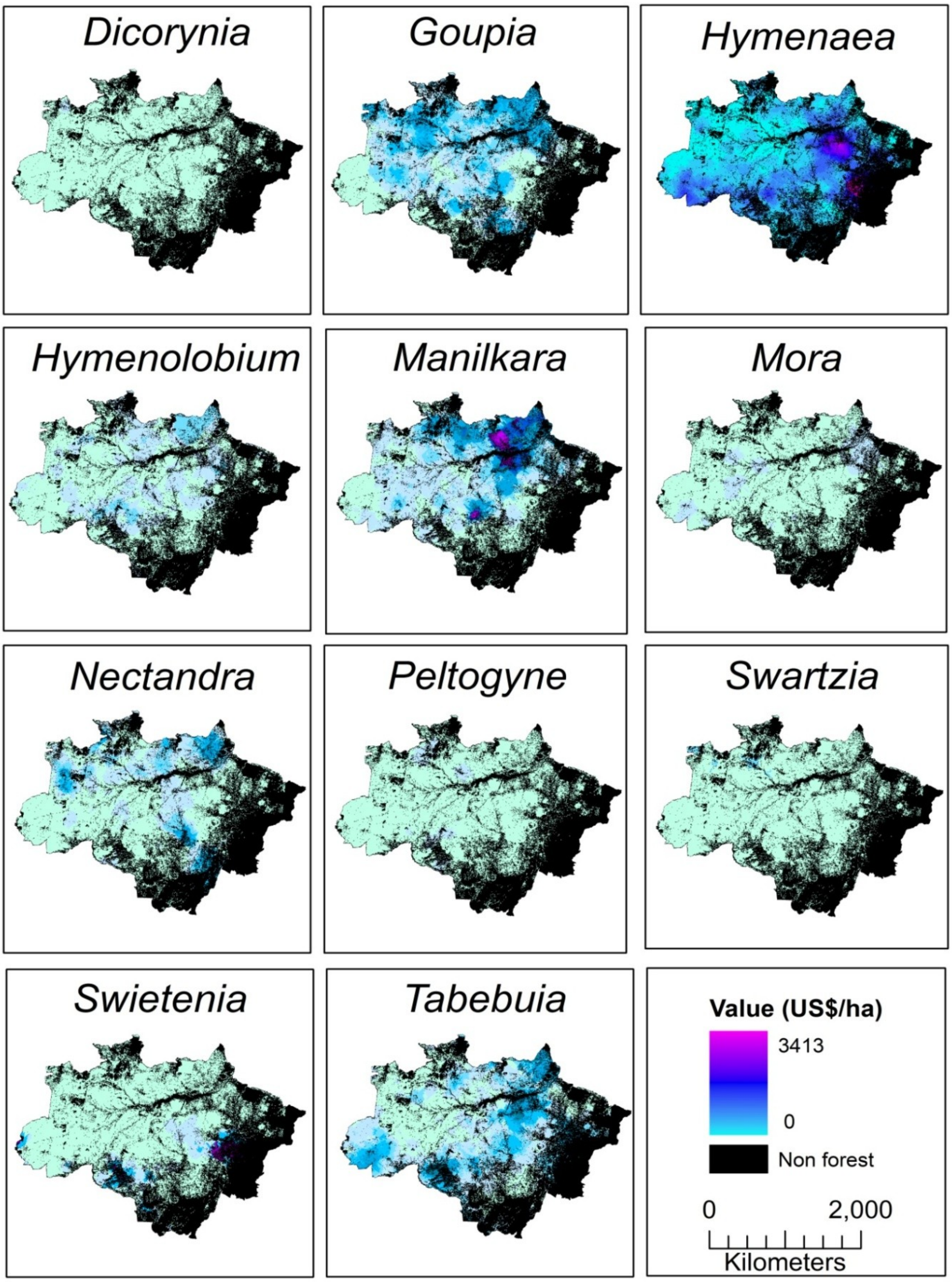


Figure 4.1. Predicted timber value across the Amazon for each of 11 high-value genera, showing genus-specific spatial patterns in the distribution of timber value and differences in total value (US$.ha^{-1}). *Hymenaea* appears to consistently have the highest value across the Amazon. Other genera, such as *Manilkara, Nectandra* and *Tabebuia* show 'hotspots' of high values in relatively restricted areas of the Amazon. Some genera, such as *Dicorynia* and *Peltogyne*, show lower values that do not vary much across space.

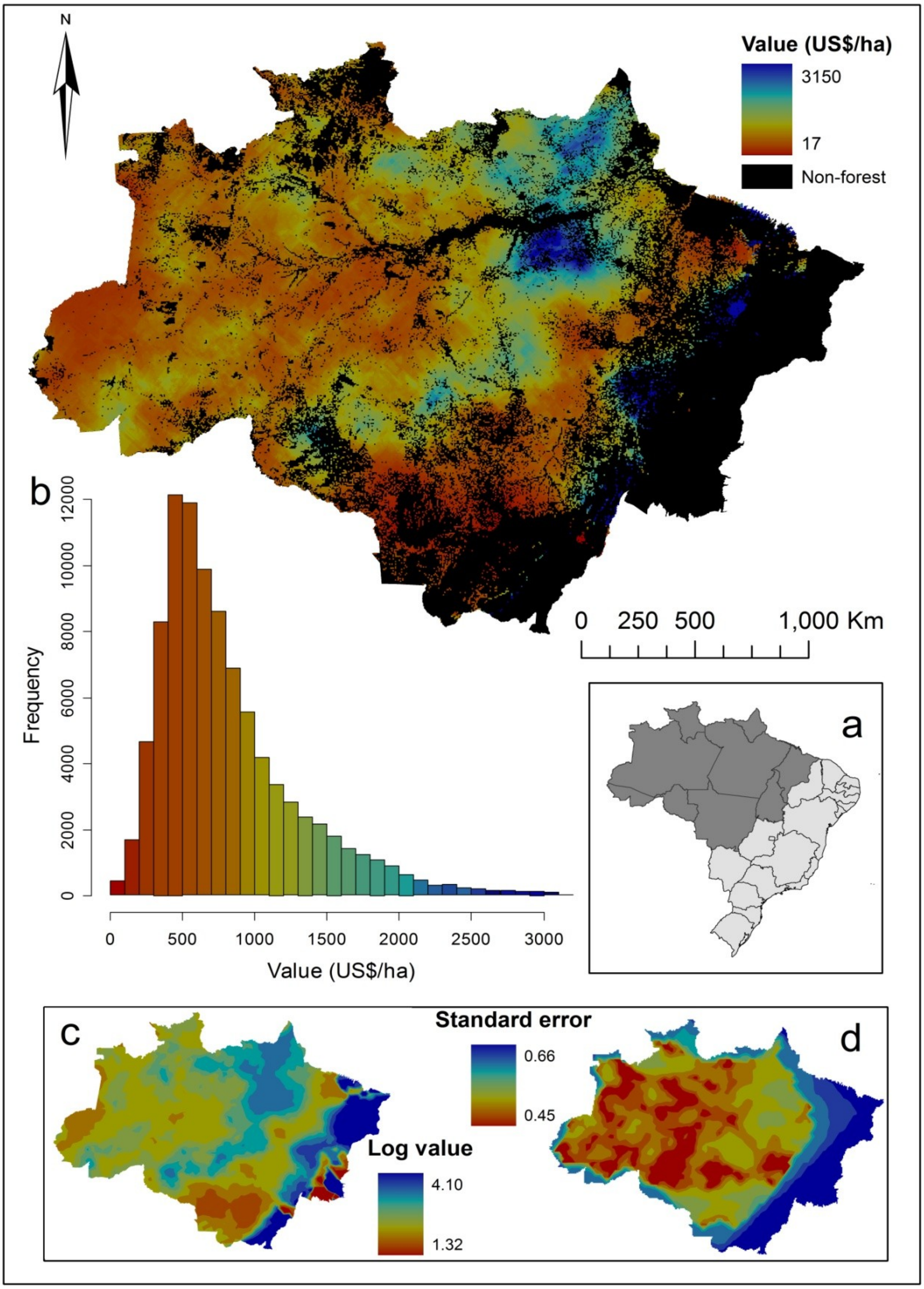


Figure 4.2. Map of timber value in the Amazon. Values range from low (US$17 per ha) to high (US$3150 per ha). a) dark shading shows the spatial extent of the Brazilian Amazon within Brazil, including the state boundaries; b) frequency distribution of timber values (US$.ha^{-1}) in the Brazilian Amazon, calculated over 151,073,784 equal-area grid squares of area 0.25 km^2; c) log- transformed timber value across the Amazon without non-forest areas

removed (the raw outputs from the kriging analysis). As non-forest areas are not removed the maximum predicted value in panel c) is 4.10, equating to US$4400 per ha, is higher than the maximum value of US$3150 per ha in the main panel. This discrepancy arises because the maximum values in panel c) occur along the eastern margins of the Amazon where, in fact, there is little forest standing. d) standard error of predicted log-transformed timber value.

Table 4.1. Mean predicted value of timber (to the nearest US$.ha^{-1}, plus 95 % confidence intervals) across the Brazilian Amazon by genus, as calculated over 151,073,784 equal-area grid squares of 0.25 km^2 (data presented in Figures 4.1 and 4.2). In some cases the lower 95% CI is the same as the mean, reflecting the heavily left-skewed kriging predictions (i.e. Figure 4.1b). The number of RADAMBRAZIL locations each genus was recorded at and the average export value (US$.m^{-3}) are also shown.

Genus	**Mean value (US$.ha^{-1})**	**Lower 95% CI (US$.ha^{-1})**	**Upper 95% CI (US$.ha^{-1})**	**Number of locations**	**Value (US$.m^{-3})**
Dicorynia	5	5	10	21	158
Goupia	221	221	587	824	257
Hymenaea	394	376	2136	945	569
Hymenolobium	18	18	26	261	96
Manilkara	208	208	1438	598	281
Mora	14	14	27	98	216
Nectandra	179	173	523	1100	292
Peltogyne	20	20	34	255	143
Swartzia	42	42	57	909	105
Swietenia	40	40	250	57	1096
Tabebuia	107	107	232	757	317
Total	813	477	4161	2465	

The 11 genera contributed very different amounts to the overall timber value, and exhibited genus-specific spatial patterns of timber value (Figure 4.1 & Table 4.1). The genus *Hymenaea* contributed the most to the total values (mean value US$394 per ha; 95 % confidence interval 376-2136), whereas *Dicorynia* contributed the least (mean value US$5 per ha 95 % confidence interval 5- 10). The mean total timber revenue predicted was US$813 per hectare (95 % confidence interval 477– 4161), with a distribution showing a few locations with particularly high values of > US$1500 (Figure 3.2). There was, however, one area known to be of relatively low value along the eastern edge of the study area that was predicted to be high value. This was likely a result of the kriging interpolation, as there were no survey locations in this area to moderate the predictions. Standard errors on the kriging interpolation (Figure 3.2d) reflect this, showing a band of high error along the eastern edge of the study area.

4.4. Discussion

The predicted timber revenue values are comparable to other estimates of timber value in the Amazon. For example, a report by Nepstad *et al.* (2007) predicts a maximum *net* value of timber in the Amazon to be US$550, whereas we predict a *gross* maximum value of US$3150 and mean of US$813 per hectare. Once the costs of converting standing timber into sawn wood are taken into account, our figures are comparable. However spatial patterns of high value timber differ between our predictions and those of Nepstad *et al.* (2007), primarily because they considered net value (i.e. profit) while we consider gross value. So, while we find the highest value areas to be concentrated in the north east, they reported that high value areas were concentrated around transport systems that offer cost effective access to the forest.

The differences among genera in their predicted values can be attributed to several factors, of which the first is variation in the spatial distributions of the genera themselves. For example, *Dicorynia* is mainly recorded in the north-west of the Brazilian Amazon and was only present in 21 of 2465 locations. Other genera, for instance *Hymenaea, Nectandra* and *Swartzia* were well represented, being present in 945, 1100 and 909 of 2465 locations respectively. They are also relatively evenly distributed across the Brazilian Amazon rather than being clustered in a small region, which influences kriging predictions. Second the value per cubic metre also varies considerably between genera, for example the average value of *Swartzia* species is US$105 per ha whereas the average value of *Tabebuia* is US$317 per ha, further influencing the different predicted values of the 11 genera. Third, differences in the abundance (i.e. volume) of each genus influenced the predicted value, for instance *Dicorynia* had a total recorded volume of just 101.24 m^3 whereas *Hymenaea* had a total recorded volume of 4528.30 m^3. As a result of these three differences among genera, when their individual values are combined the genus-specific patterns are masked and the total values across the Amazon average out. Thus, simply adding the 11 separate genus level kriging predictions does not produce the same results as kriging the total value directly. This averaging effect explains why the maximum genus-specific value obtained of US$3143 per ha is slightly higher than the maximum total value which is US$3150 per ha.

There is a clear spatial pattern in timber value across the Amazon (Figure 4.2), with the most valuable timber in the north eastern region. Various studies have established that there is an east to west gradient in average wood density/wood specific gravity, with high wood densities occurring in the east (Baker *et al.* 2004, Chave *et al.* 2006, ter Steege *et al.* 2006, Baker *et al.* 2009). This pattern in wood density is also associated with gradients of increasing seed mass in the east (ter Steege *et al.* 2006), higher above ground live biomass (AGLB) in the northeast and central Amazon (Saatchi *et al.* 2007), and a threefold variation in coarse wood production with higher production in the west (Baker *et al.* 2009, Malhi *et al.* 2004).

The northeast region of the Amazon has old, nutrient poor, well drained soils and a moderately seasonal climate that is occasionally subject to drought (Malhi *et al.* 2002). By contrast, the less valuable areas sit in the west and to the south, where there are richer soils, a more seasonal climate and a more dynamic environment in terms of individual tree turnover (ter Steege *et al.* 2006, Malhi *et al.* 2002). These edaphic (soil) and climatic conditions help to explain the emergent pattern of timber value we have presented. ter Steege *et al.* (2006) identified two primary gradients in tree composition in the Amazon; the first gradient parallels a major gradient in soil fertility, and the second composition gradient is related to climate, specifically dry season length. They found that in the east the most abundant genera are legumes, yet none of the most abundant genera in the western Amazon are legumes. Thus, unsurprisingly, the poor soils of the east appear then to favour species that are able to cope with low nutrient levels, these species tend to be long-lived trees with slow growth rates but high wood density. Conversely, the more fertile soils in the west are associated with higher growth rates (Malhi *et al.* 2004), lower wood densities (Baker *et al.* 2004), high productivity and a high turnover of individuals (Phillips *et al.* 2004, ter Steege *et al.* 2006), with an average stem turnover rate of 2.6 $\%.yr^{-1}$ in the west/south-west compared to a rate of just 1.35 $\%.yr^{-1}$ in the northeast (Stephenson & Van Mantgem 2005, Quesada *et al.* 2009). Additionally, ENSO (El-Nino Southern Oscillation) causes episodic droughts in the eastern and central Amazon (Malhi & Wright 2004), however it has little affect on rainfall in the south-west. High wood density species often have a lower vulnerability to drought stress (Chave *et al.* 2006) and are thus less affected by episodic drought than their faster growing light wood counterparts. Mean stand level wood specific gravity was found to be 15.8 % higher in eastern and central Amazonia compared to western Amazonia (Baker *et al.* 2004). Given that high value trees often have a high wood density and take a long time to grow (mean wood density is inversely correlated with wood productivity (Malhi *et al.* 2006)), the highly productive, high turnover areas of the west are not ideal for slow growth trees. However, the poor soils of the east that competitively favour legumes (family, *Fabaceae*), which includes seven of the 11 valuable genera considered in this study (*Swietenia, Mora, Swartzia, Peltogyne, Hymenolobium, Hymenaea, Dicorynia),* are more suited to slow growing, stress resistant species that are out competed on richer soils.

We were not able to include all economically valuable timber genera in this study, omitting genera such as *Carapa* that have high average prices. It was necessary to select genera which were present both in the RADAMBRAZIL data set and for which export value (US.$$m^{-3}$) data was available. It is reasonable to assume that inclusion of other economically important timber genera could alter the absolute values emerging from our analyses. However, given the ecological similarities among many economically important timber trees, we feel that the general spatial patterns would remain the same, with the highest value tree stands being located in the northeastern region of the Brazilian Amazon. Another point to note is the possibility that allied genera were confused in the field surveys, with genera such as *Swartzia* and *Bocoa* sometimes misidentified. However, we again feel that the general trends found here would be robust to minor errors in the field data, partly because no single genera contributes more than 19.66 % of all individuals analysed.

We have shown that the patterns of standing timber value in the Amazon reflect known, large-scale ecological gradients extending across the Amazon, determining the spatial distribution of functional traits of trees which are, in turn, correlated with timber values. We expect that understanding the spatial patterns of timber value across the Amazon will aid predictions of logging movements and thus predictions of potential future road developments. These predictions in turn will be of great use in estimating the spatial patterns of deforestation in this globally important biome.

Chapter 5: **Temporal patterns of road network development in the Brazilian Amazon**

5.1. Abstract

The Brazilian Amazon is a globally important ecosystem that is undergoing rapid development and land use change. Roads are a key spatial determinant of land use conversion and strongly influence the rates and patterns of habitat loss, and represent a key component of models that attempt to predict the spatio-temporal patterns of Amazonian land use change and the consequences of such changes. However, the spatio-temporal patterns of road network development are poorly understood and seldom quantified. Here, we used manually digitised satellite imagery at multiple temporal and spatial scales across the Brazilian Amazon to quantify and model the rate at which road networks are proliferating. We found that the road network grew by almost 17 000 km per year between 2004 and 2007. There was large spatial variation in road network density, with some municipalities having road densities as high as 0.5 km/km^2, and road network growth rates were highest in municipalities with an intermediate road network density. Simulations indicated that road network development within municipalities follows a logistic growth pattern through time, with most of the development occurring within a 38 year time period. This time period is similar to those of other boom and bust development dynamics observed in the Brazilian Amazon. Understanding the temporal patterns of road development will aid the development of better predictive land-use change models for the Amazon, given the key importance of roads as a predictor of deforestation in many existing models.

5.2. Introduction

Global road networks have been expanding at a rapid rate since the 1900's (Forman *et al.* 2003). Roads are a distinctive feature in any landscape, with many countries giving 1-2% of their land surface over to roads and roadsides (Forman 1998). However, the ecological effects of roads spread beyond the physical footprint of the network and may impact 15-20% of the land or more (Forman & Alexander 1998). In the context of tropical deforestation, roads cause a relatively small amount of direct habitat loss, but exert a huge indirect influence on spatial patterns of deforestation by allowing easier access to new frontiers (Fearnside 2008, Geist & Lambin 2002, Perz *et al.* 2007, Perz *et al.* 2008). Roads also encourage extractive industries and further deforestation by settlers, thereby indirectly influencing deforestation rates. As roads are developed to access resources, roads may be seen as a cause of development and it is this causal relationship that led to the Brazilian development policies of the 1970's based around road construction (Alves 2002, Kirby *et al.* 2006). However, roads may also be a consequence of development, where, as economies grow they require better infrastructure support and so road networks are developed.

The influence of roads on spatial patterns of deforestation ensures they also exert a strong influence of spatial patterns of biodiversity loss (Forman 1998, Maki *et al.* 2001, Fearnside 2005, Finer *et al.* 2008). Roads influence biodiversity directly through road kill events, but again the largest impact of roads is through indirect processes and 'extended effects'. For example, roads can alter abiotic processes such as surface run off (Forman & Alexander 1998) and microclimatic conditions including light levels, air and soil temperature, air and soil moisture, soil pH and nutrient levels (Gehlhausen *et al.* 2000, Delgado *et al.* 2007, Honu & Gibson 2006). Roads fragment forest habitats, creating new habitat edges and acting as barriers to the movement of animals (Richardson *et al.* 1997, Dyer *et al.* 2002, Arima *et al.* 2005, Keller & Largiader 2003, Goosem 2007, Rico *et al.* 2007, McGregor *et al.* 2008). These extended effects alter the abundance, distribution and behaviour of species over large areas (Vos & Chardon 1998, Blom *et al.* 2005, Potvin *et al.* 2005, Bee & Swanson 2007, Eigenbrod *et al.* 2008). Some species do appear to prosper with the presence of roads, benefitting from additional resources such as road killed carrion (Rydell 1992, Laurian *et al.* 2008). However, negative effects are five times more prevalent than positive effects (Fahrig & Rytwinski 2009).

The Brazilian Amazon contains approximately one third of the world's remaining rainforest, covering an area of 4.1 million km^2. The region is highly biodiverse with 10-20 percent of the planet's known species, is one of the three most bioculturally diverse areas in the world (Loh& Harmon 2005), and provides many valuable ecosystem services such as water regulation (Feanside 2005, Foley at al. 2007, Bradshaw *et al.* 2007), carbon sequestration (Asner *et al.* 2004a, Foley at al. 2007), and local and global climate regulation (Foley at al. 2007, Bonan 2008, Malhi *et al.* 2008). However, the Brazilian Amazon is also rapidly undergoing extensive development with widespread land-use conversion. Roads are a key spatial determinant of land use conversion in this region, dictating the spatial pattern of deforestation by regulating access to standing forests which are logged for timber and later clear-cut to make way for agriculture or pasture, which secures land ownership via 'productive use' (Fearnside 2005, Kirby *et al.* 2006, Perz *et al.* 2008). This process of deforestation and agriculture following roads has been well documented (Geist & Lambin 2002, Walker *et al.* 2004, Perz *et al.* 2007, Fearnside 2008). Given that roads are a key spatial determinant of land use conversion and that they have extensive impacts on rates and patterns of habitat loss, it is important that we know how many roads are being built, how fast road networks are developing and where they are developing in this globally important ecosystem.

Many studies have investigated temporal patterns of land use change in the Amazon (e.g. Dale *et al.* 1994, de Koning *et al.* 1999, Carpentier *et al.* 2000, Soares-Filho *et al.* 2002, Walker *et al.* 2004, de Barros Ferraz *et al.* 2005, Walsh *et al.* 2008, Wassenaar *et al.* 2007, Moreira *et al.* 2009, Araujo *et al.* 2009, Muller *et al.* 2010), as well as the temporal patterns of change in the forces that drive land use change such as economic and agricultural trends (Morton *et al.* 2006, Ewers *et al.* 2008, Araujo *et al.* 2009). However, knowledge of the temporal dynamics of road network development lags far behind, with just one study having examined this in the Brazilian Amazonian state of Pará; Brandão & Souza (2006) mapped a total of 25 196 km of roads, of which 15 727 km were constructed in a 10 year time period (1991-2001). This stands in stark contrast to the acknowledged importance of temporal changes in road networks for understanding and predicting deforestation rates and patterns. For example, at least half of the land use change models based in the Amazon that we are aware of use road networks as either a spatial or temporal predictor of deforestation (Laurance *et al.* 2001, Messina & Walsh 2001, Walker *et al.* 2004, Soares-Filho *et al.* 2006, Etter *et al.* 2006, Michalski *et al.* 2008, Araujo *et al.* 2009, Mena *et al.* 2011). In Central Africa, Laporte *et al.* (2007) showed that rates of road building in Central Africa increased over 31 years as a function of industrial logging. Yet no similar study has been carried out for the Amazon or any of its nine constituent countries despite the Amazon being at the forefront of global tropical deforestation.

Much of the existing literature on road network development in the Brazilian Amazon and elsewhere is qualitative rather than quantitative. For example, Taaffe *et al.* (1963) described an idealised progression of transport development using Ghana and Nigeria as examples. Perz *et al.* (2007) investigated the differing development histories of four road networks through interviews with local residents, and qualitatively compared the social and spatial processes behind different patterns of unofficial road development. More rigorously, Maki *et al.* (2001) documented the development of a road connecting two urban centres in Peru (Iquitos and Nauta), and Brandão & Souza (2006) documented the growth rate of roads for an area of 546 000 km^2 in the state of Pará (Brazil), finding that the road network nearly doubled over a period of 10 years. There have also been some recent attempts to model the paths that individual roads take as they are constructed (Arima *et al.* 2008). Yet there have been no attempts made to understand how rapidly road networks are developed over large spatial scales, despite that knowledge being fundamental to understanding land use change trajectories in the region (Barlow *et al.* 2011).

Here, we used pre-existing road maps, satellite imagery and simulations to document the historical, and predict the future, temporal dynamics of road network development in the Brazilian Amazon. We conducted our analysis at two spatial and temporal scales. First, we generated annual road maps for a nine year period (2000-2008) from areas of low, medium and high road density to investigate annual patterns of network growth at relatively small spatial and temporal scales. Second, we used pre-existing road maps for the entire Brazilian Amazon from 2004 and 2007 to investigate rates of road network growth within municipalities. Based on our analyses of observed spatial and temporal variation in road network development, we constructed simulations to determine the temporal trajectory of road network growth in this region.

5.3. Methods

We manually digitised annual Landsat 5 TM images from 2000 to 2008 at each of three locations (path/row ID's: 231/065, 227/066, 227/062). Due to the nature of the method we were unable to distinguish between road types (e.g. official, unofficial or paved, unpaved), thus 'roads' refer to all road types aggregated together. Landsat road densities were divided in ArcGIS into three density classes based on natural 'Jenks' breaks, low (0.00-0.05 km/km^2), medium (0.05-0.13 km/km^2) and high density (0.13-0.23 km/km^2) scenes. From within these classes three Landsat scenes were selected to represent regions of low (0.02 km/km^2), medium (0.07 km/km^2) and high (0.14 km/km^2) road density (Figure 5.1), in the expectation that regions with different road density are likely to have different trajectories of road network development. Although digitisation and analysis was carried out at the Landsat scene scale a similar pattern of road densities were observed at the municipality scale (Figure 5.1). There are many automated approaches to digitising road networks (Mena 2003; Brandão & Souza 2006, Li & Briggs 2009, Movaghati *et al.* 2010), but these are typically less accurate than manual digitisation (Li & Briggs 2009) in which images are observed and visible roads are traced individually and by hand. Subsequently, we manually digitised images in ArcGIS, following the methods described by Brandão & Souza (2006) to create road maps for each year at each location. This method was validated with ground truthing by Brandão & Souza (2006). In addition to the methods detailed by Brandão& Souza (2006), we created the temporal series of maps sequentially based on the previous year's map. This meant that any gaps in the imagery (e.g. obscured by cloud/canopy cover) in one year would be recorded the previous or next year. We validated our digitised satellite images for 2007 against a road map of the whole Amazon from 2007 that was produced using the same methods outlined by Brandão & Souza (2006) (map data provided by IMAZON), by determining the spatial congruence, i.e. how much of the digitised roads overlapped between the two maps. We found that in terms of total roads digitised there was a good similarity between the two maps at all three locations (low density= 92%, mid density =79%, and high density 98% similarity). Across the three locations, 84% of our digitised roads fell within 200m of the Amazon map, and 81% of the roads on the Amazon map fell within 200m of our digitised roads. Much of the variation in manually digitised maps arises from variation in judgement calls on irresolute roads, for example while one person may decide a faint line is a road another may decide it is a boundary line between two patches of land and not a road. Thus while a spatial congruence of 100% is not expected, our average congruence of 82.5% suggests a good degree of accuracy in identifying roads with this method.

To examine how road networks changeover time and if changes varied in different density classes, we calculated the cumulative amount of road (km) between 2000 and 2008 for each Landsat image. Changes in road length was analysed as a function of year and location using an ANCOVA in the statistics program R 2.10.1 (R Development Core Team 2009).

To investigate the spatio-temporal patterns of road network development at the scale of the entire Brazilian Amazon, we used two complete road maps for the region that were generated using the same manual digitisation methods on Landsat imagery taken in 2004 and

2007 (Brandão & Souza 2006). We used 443 municipalities of the Brazilian Amazon as spatial units for analysis, determining the initial road density in 2004 (km/km^2) and the change in road density between 2004 and 2007 (km/km^2) for each municipality. We chose to carry out analyses at the municipality level because this is the level at which decisions pertaining to development are made.

Using multiple regression we tested the effect of the official road density and various socio-economic factors had on road network growth rates. Census data by municipality was obtained for the year 2004 from IPEA (IPEA 2012) for; permanent agricultural area, temporary agricultural area, total agriculture area, cattle head count, credit available for agriculture, credit available for cattle. The percentage of land area protected was also included. Rural and urban population counts for 2000 were used because data for 2004 were unavailable. Data were log-transformed where appropriate and agricultural area measures were converted to density to control for differences in municipality area. Starting with a complete model including all variables, we used model simplification to identify factors relevant to road network growth rates. We also tested for an effect of initial road density on road network growth rates using linear regression, including a quadratic term to allow for a non-linear relationship. Preliminary analyses showed that these models explained little of the variance in the data (R^2=0.06), but visual inspection of figures suggested that maximum rates of road network development varied with initial road density and therefore that initial road density might act as a limiting factor on road network growth rates. The effect of a limiting factor can be quantified using quantile regression (Cade *et al.*1999), so we estimated how initial density limits road network growth rates using quantile regression as implemented by the R package ‘Quantreg’ (Koenker 2010).

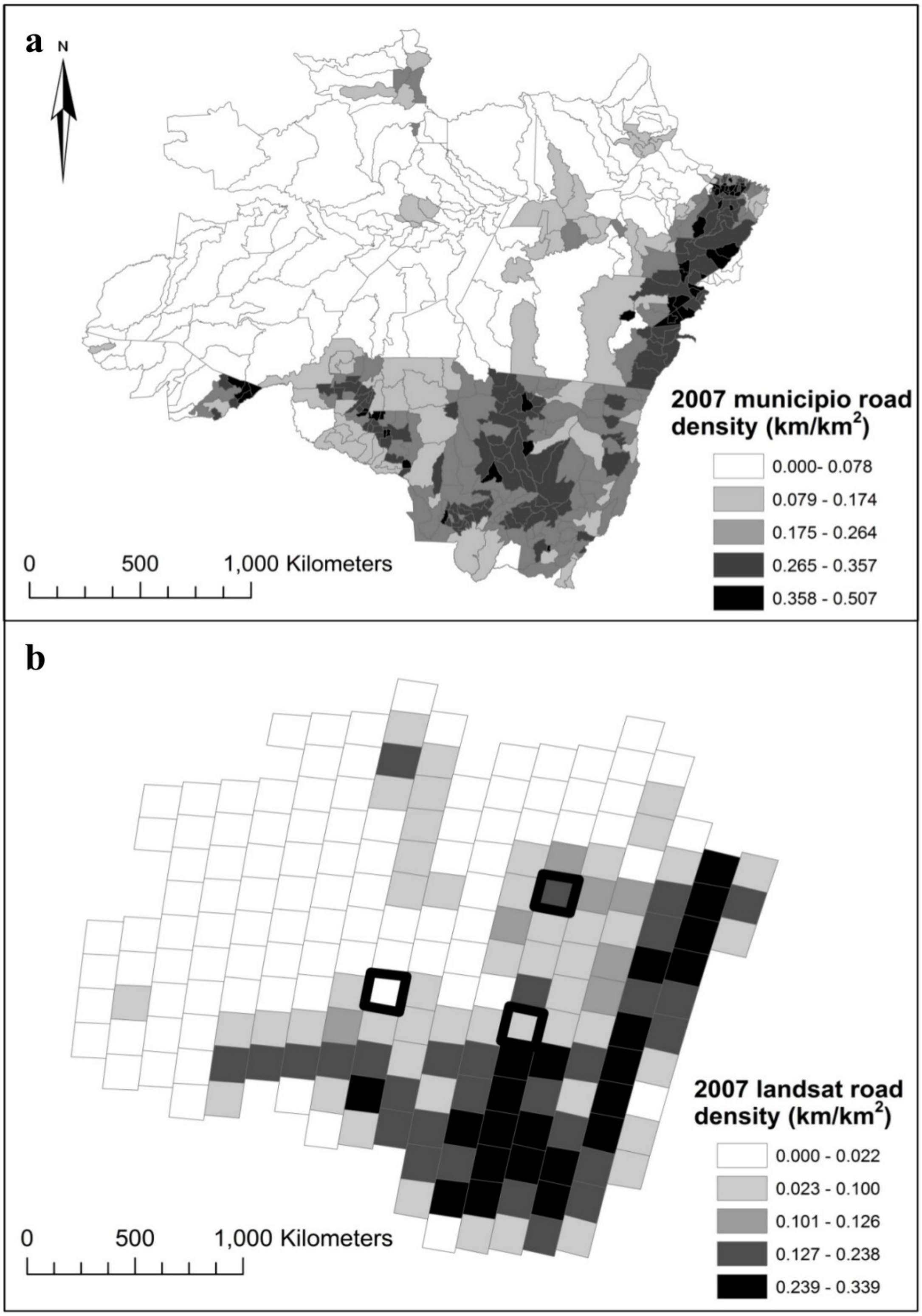


Figure 5.1. Spatial patterns of road network density in the Brazilian Amazon for 2007. Spatial units represent a) municipality boundaries and b) at Landsat scenes. The three dark squares show the three Landsat scenes in which we analysed annual changes in the road network

To investigate the likely temporal trajectory of road network development in the Brazilian Amazon, we constructed a simulation (essentially a Markov-chain) based on the relationship we observed between initial road density and road network growth rate. The simulation was designed to estimate the length of time it would take for a municipality to develop a road network to such a density that new roads are no longer being created. The maximum road density observed in any municipality in 2004 was 0.5 km/km^2, this municipality showed no development in 2007, and so we assumed that 0.5km/km^2 is the density at which road development stops. We assumed that municipalities would have an initial road density of 0.0 km/km^2. In each annual time step of the simulation, we determined the amount of new road that would be developed in that municipality by creating a window (data bin) around the observed initial density of ±0.05 km/km^2 and randomly selecting one observed growth rate that occurred within that window. For example, given an initial density of 0 km/km^{2}the window will subset out all growth rates corresponding to municipalities with densities of between 0 and 0.05 km/km^2. A random growth rate is selected from the subset rates (e.g.0.03 km/km^2) and is added to 0km/km^2, giving the initial density for the next time step (i.e. 0.03 km/km^2). All growth rates within a subset have an equal probability of being selected. This method takes into account the non-uniform distribution of growth rates observed among municipalities with similar initial road densities and allows for 'kick-start' development, as observed in some municipalities that had very low initial road network densities but very high rates of road network growth. The road network density in a simulation progressively increased through time, and we stopped all simulations when a road density of 0.5 km/km^2 was reached. For each of 1,000 simulations, we recorded the cumulative increase in road density over time, and fitted a logistic model to the simulated data using the R package 'grofit' (Kahm *et al.* 2010). We used the logistic model to quantify the length of the lag and boom phases of road network development. The lag phase tells us the number of years a municipality might take from initial colonisation until it experiences an exponential increase in road network growth. The boom phase, in turn, begins immediately at the end of the lag phase and represents the number of years during which road networks are rapidly growing. We used the second derivative of the logistic model to determine the start and end point of the boom phase. The local maxima and minima points on the second derivative of a logistic model indicate the two time points at which the rate of change in the road network is at its greatest (Ewers &Didham 2006), either accelerating at the end of the lag phase or decelerating at the end of the boom phase.

5.4. Results

Road network development showed significant spatial variation across the Amazon (Figure 5.2). While some municipalities experienced no network development, others experienced development as high as 0.064 km/km^2/year. Interestingly, most road network development was concentrated in the municipalities that form the 'Arc of Deforestation' along the south and eastern edges of the Amazon (Fearnside 2005). There was little development ahead of the Arc of Deforestation and almost none behind it to the south.

Significant amounts of new road were added to the road networks in all three Landsat scenes analysed, with road construction occurring in each year × scene combination (Figures, 5.2 & 5.3). Road network growth rates were fairly constant through time at all three density scene areas (R^2=0.98, $p<0.001$, df= 21, Figure 5.3), although in the high density scene there was a large change in density between 2000 and 2001. The average rate of increase was lowest in the low density Landsat scene (mean growth rate 76.9 km/km^2/year ±37.3 SE), highest in the mid density scene (289.7 ±37.3 km/km^2/year) and intermediate in the high density scene

(275.4 ±26.4 km/km^2/year). These differences in road network growth rate were significantly different between the low and mid density scenes (p<0.001) and between low and high density scenes (p<0.001), but not between the mid and high density scenes (p=0.705, Table 5.1).

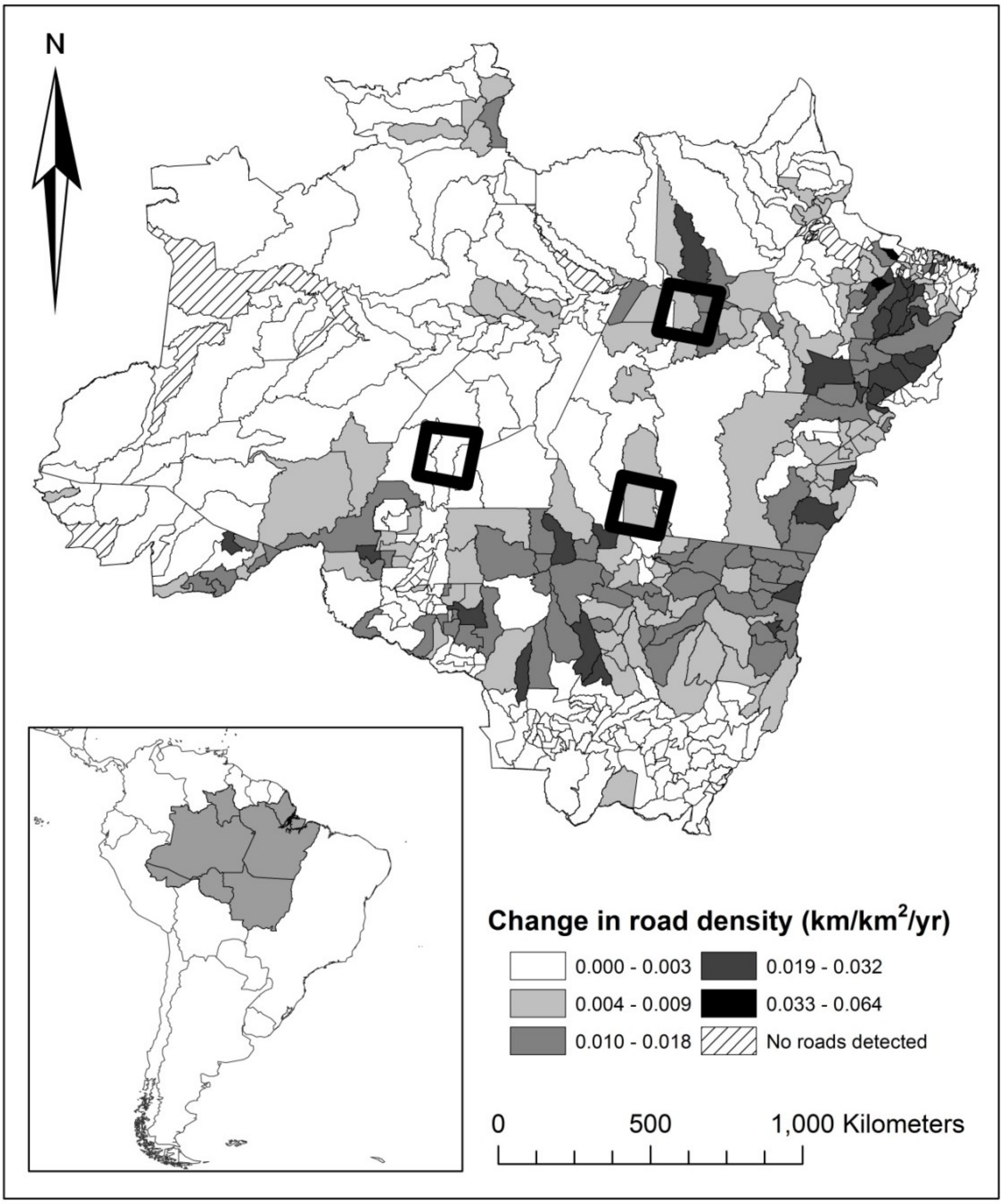


Figure 5.2. Spatial patterns of road network growth rates in the Brazilian Amazon. Spatial units represent municipality boundaries.

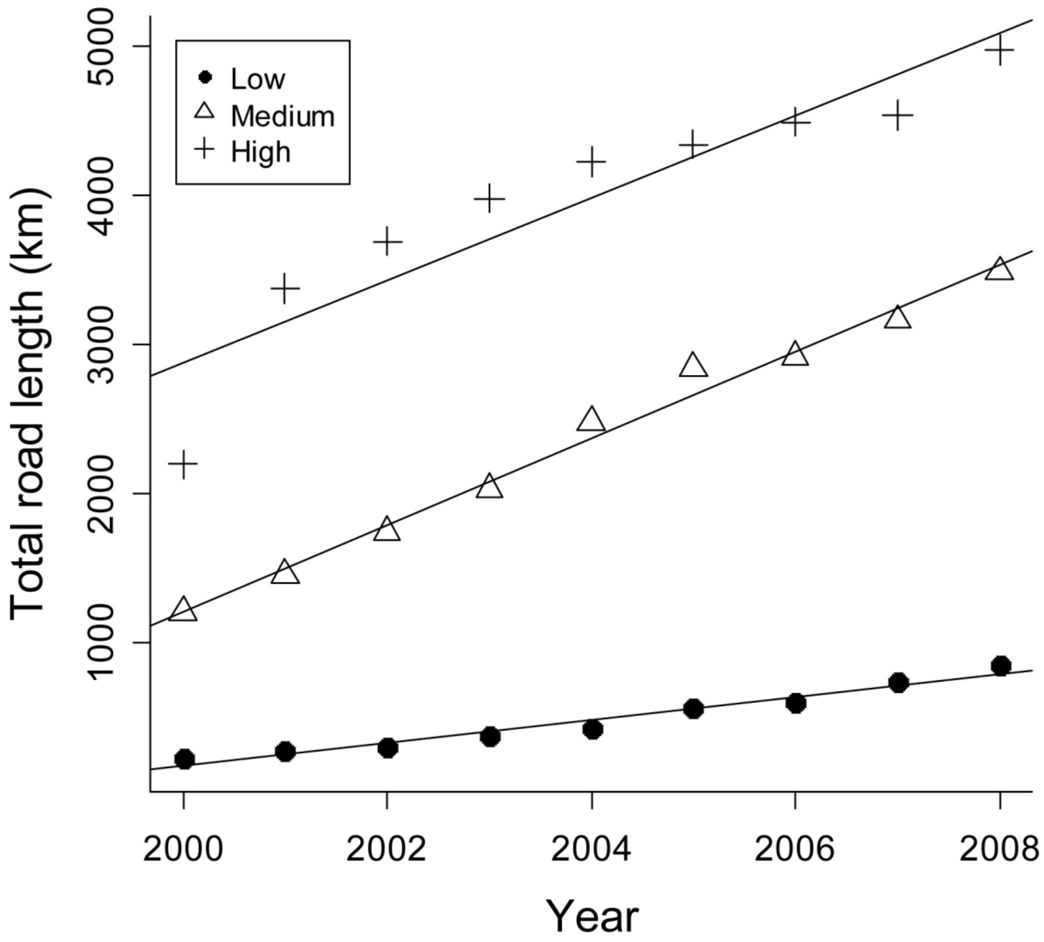


Figure 5.3. Annual cumulative growth in the road network in three regions of the Brazilian Amazon. The three regions correspond to the Landsat scenes shown in Fig 1 and varied in the degree of initial road density: low (0.02 km/km^2); medium (0.07 km/km^2; and high (0.14 km/km^2)

Table 5.1. ANCOVA comparing differences in road development, in terms of total length (km) in three Landsat scene locations over a nine year period (2000-2008).

Density scene	Slope	Standard error	t value	P value	d.f.	R^2
High	275.39	26.43	10.42	<0.001	21	0.98
Low	76.90	37.38	-5.31	<0.001		
Medium	289.7	37.38	0.38	0.705		

At the scale of the Brazilian Amazon, 50 666 km of new roads were constructed between 2004 and 2007 (Figure 5.2). We found that change in road density between 2004 and 2007 was significantly negatively related to official road density, permanent agriculture density in 2004 and protected area percentage, however was significantly positively related to credit available for agriculture (Table 5.2). Further, there was a significant interaction between official road density and permanent agricultural density, which was significantly positively related to change in road density. Despite the significant relationships observed, the amount of variance in change in road density explained was low, just 18% ($p<0.001$, $F_{5,437}=20.5$, d.f.= 437, $r^2= 0.18$). Other socio-economic variables tested showed no significant relationship with change in road density and were removed during model simplification.

Table 5.2. The effect of socio-economic factors and official road density on the change in road density between 2004 and 2007, across the Brazilian Amazon.

Factor	Slope	Standard error slope	t value	P value
Log official road density	-0.223	0.054	-4.12	<0.001
Log permanent agriculture density	-0.024	0.005	-5.22	<0.001
Log credit available for agriculture	0.001	0.0001	4.56	<0.001
Log protected area percent	-0.001	0.001	-2.13	<0.05
Log official road density interaction with log permanent agriculture	0.621	0.072	8.65	<0.001

We found a weak, but significant, relationship between initial road density and change in density over three years using linear regression with a quadratic term ($p<0.001$, $F_{2,440}=15.28$, $r^2=0.06$). Quantile regression revealed that for any given initial road network density there was a maximum potential increase in road density (90th percentile regression, $p< 0.001$, $F_{7,435}=29.30$). Many municipalities had road networks that did not grow between 2004 and 2007, regardless of initial road density, whereas the most rapid road network growth rates occurred in municipalities with intermediate road densities (Figure 5.4).

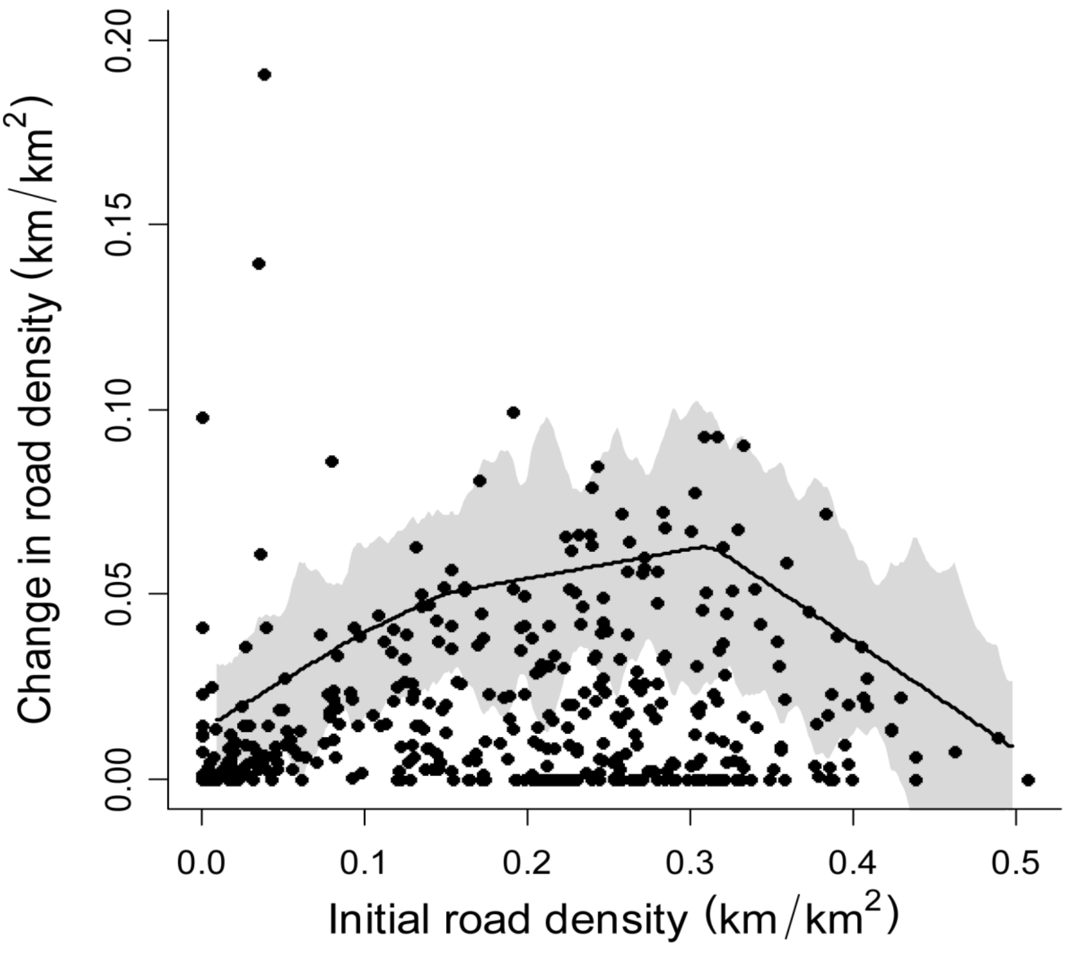


Figure 5.4. Relationship between initial road density on change in road network density over a three year period (2004 to 2007) in 443 municipalities of the Brazilian Amazon. The black line shows the 90th percentile regression, and represents the maximum likely road network growth rate for a municipality with a given road density. The grey shaded area represents the 95% confidence interval around the 90th percentile regression

Our simulation found that it takes an average of 75 years (95% CI 58-103 years) for a municipality to fully develop a road network (Figure 5.5). The lag phase of road network development typically lasted 15 years (95% CI 1-34 years), after which the boom period lasted an average of just 39 years (95% CI 19-71 years).

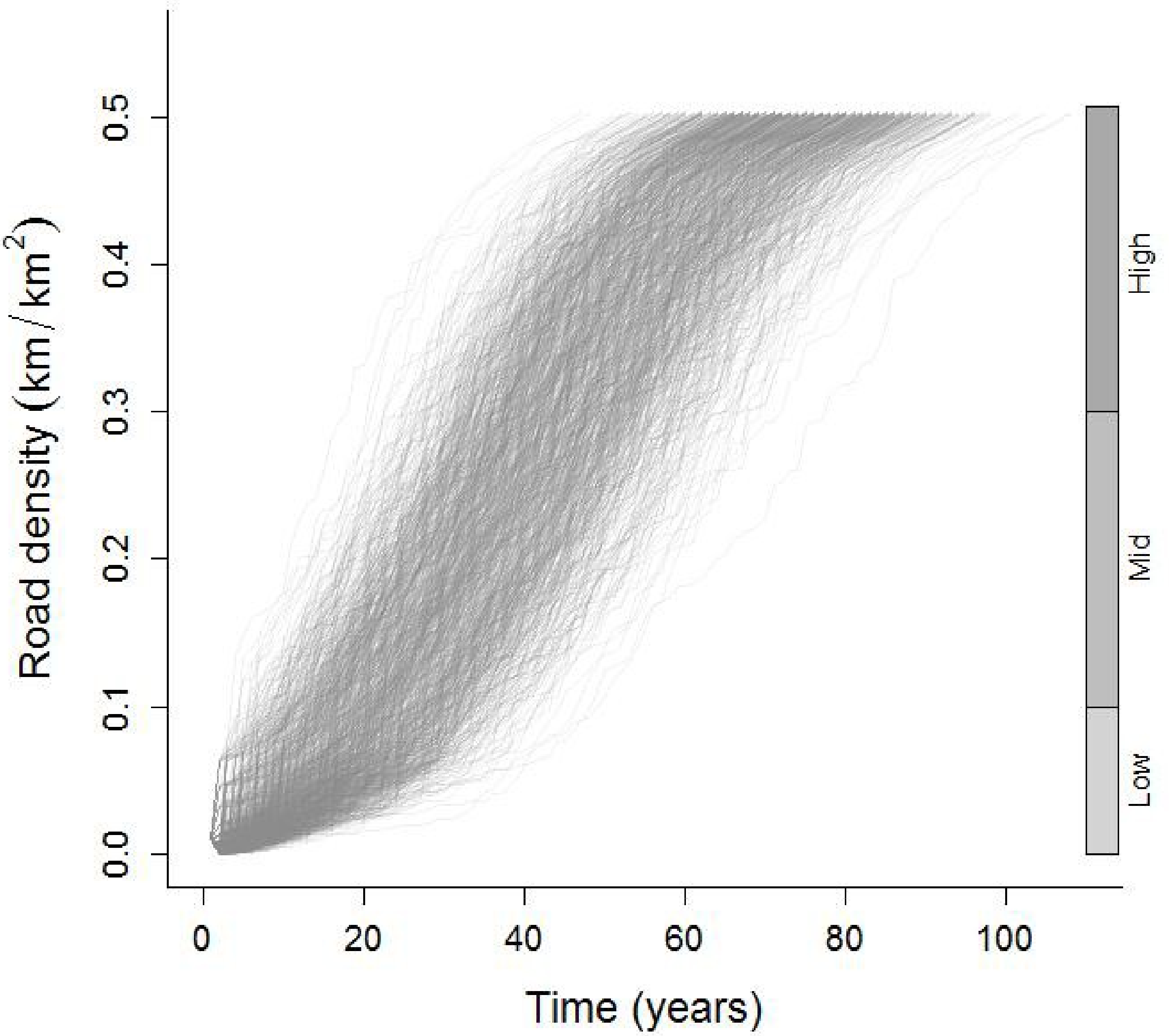


Figure 5.5. Simulated trajectories of road network development in the Brazilian Amazon. Each line represents the trace from one of 1000 simulations, estimating the road density within a municipality as a function of municipality road network age. The bar shows the delineation of low, mid and high density municipalities. Delineation is based on natural 'Jenks' breaks.

5.5. Discussion

We found that, on average, 16 889 km of new roads are added every year to the Brazilian Amazon road network, and that road network development within municipalities is a process that likely spans around four decades. Road network growth rates varied considerably among municipalities, likely reflecting the different economic and developmental histories of those regions, and the simulated patterns of road network development indicated it follows a boom and bust dynamic in which the majority of roads are built in a short period of time.

There is an evident spatial pattern of road density observed, with high road densities observed in the eastern and southern Amazon (Figure 5.1), particularly in the states of Rondonia, Mato Grosso and Pará. This pattern is likely associated with settlement and economic development of the municipalities within these states. In the 1970's the Brazilian government instigated several policies to encourage the development and colonisation of the Amazon. Government incentives to encourage colonisation and cattle ranching projects (Kirby *et al.* 2006), as well as private colonisation projects via companies and co-operatives, were utilised to encourage settlement particularly in the frontier states of MatoGrosso, Rondonia and Pará (Jepson 2006). At the centre of the colonisation initiatives was the building of road networks, notably the Cuiaba-Porto Velho (BR-364), Transamazonian (BR-230) and Cuiaba-Santarem (BR-174) highways (Alves 2001, Kirby *et al.* 2006). Between 1970 and 1990 almost 700,000 families were relocated to the Amazon (Ludewigs *et al.* 2009) and it is unsurprising that road densities were highest in these states.

Within a given Landsat scene, we found that the road network growth rate was relatively constant through time over the nine year period during which we tracked annual changes to the road network. There are two main causes of road network development; government initiatives to link populations and economies, and private sector development to aid the extraction of natural resources (Brandão & Souza 2006, Perz *et al.* 2007). Both causes have considerable momentum behind them, in that constructing roads require expensive machinery and personnel. If these resources are both limiting and used with the same level of efficiency among years, it follows that the rate of road construction in any given year will be similar to the rate the year before. This potentially explains why networks grew at remarkably constant rates within Landsat scenes, and the same momentum suggests that the rates of road network development we observed are likely to continue into the immediate future.

While road networks appear to grow at relatively stable rates within a region, we detected significant differences in road network growth rates among regions. Average growth rates were low in regions with low initial road density, increased in regions with medium road density and then reduced again in regions with high initial road density. This general pattern was defined largely by an apparent upper limit to road network growth rates at any given initial road density, with the most rapid rates of network growth occurring at intermediate road densities. By contrast, at all initial road densities there were municipalities that experienced no road network development. Across the Brazilian Amazon, road networks were rapidly expanding in the set of municipalities that collectively form the 'Arc of Deforestation' along the south and eastern edges of the Amazon (Fearnside 2005). This is where most deforestation is concentrated in the Brazilian Amazon, and the level of congruence between road network growth rates and the Arc of Deforestation further reinforces the link between roads and deforestation (Fearnside 2005, Kirby *et al.* 2006).

A very small number of municipalities with very low initial road densities had road network growth rates that were much higher than the general pattern observed in the remaining municipalities (Figure 5.4). Similarly, we detected a rapid jump in road construction between 2000 and 2001 in the high density Landsat scene (Figure 5.3). These apparently sporadic events of rapid road construction at particular places and times suggests that the general pattern of road network development can be accelerated under exceptional circumstances, perhaps through significant government investment in roads as part of regional development initiatives devised to kick-start economic development (Carvalho *et al.* 2002), or possibly through the establishment of new industries. The number of these exceptions was, however, very low, indicating that there is a general pattern by which road networks develop.

The observed pattern of road network development in the Brazilian Amazon likely reflects the economic trajectory of municipalities. The first new roads created in a municipality increase access to the region for extractive industries and colonisers, who further expand the road network by constructing unofficial roads to move their products to markets (Fearnside 2008, Geist & Lambin 2002, Perz *et al.* 2007, Perz *et al.* 2008). These unofficial roads are financed by the profits arising from exploiting resources and land in a previously unexploited region. This process of increasing economic returns within a municipality can stimulate further development in the region, generating a positive feedback loop that leads initially to exponential growth and maximum network growth rates as resource extraction and land development reaches the highest levels the region can support. Once resources begin to be depleted and the majority of the land had either had its resources removed or has been converted to other uses such as agriculture, the rate of road network growth might slow for one of two reasons. First, the economic impetus to expand and develop might be reduced, leading to slower road network growth rates. Alternatively, the road network might have approached a density that is high enough to provide access to all parts of the municipality. In this case, a sufficient road infrastructure already exists and would render further expansion unnecessary. This likely explains why the majority of observed road development is concentrated along the arc of deforestation (Figure 5.2); new roads are developed in front of the arc to aid timber extraction and once the 'arc' passes the rate of new road construction will slow and eventually stop as resources are exhausted and road networks are developed to such an extent that the entire municipality is well-connected.

The relationship between initial road density and change in road density (Figure 5.4) provided a good relationship upon which to model the temporal trajectory of road development. The official road network has been used in the past to predict future road networks for use in land use land cover change modelling (Messina & Walsh 2001, Soares-Filho *et al.* 2006). Here we found that the density of official roads to be inversely related to change in road density, likely because the official road network has not changed much over the years and much of the development associated with official roads has already taken place. Thus while areas with high official road densities are likely to be areas of high road density in general, they are unlikely to be areas of high changes in road density. Arable agriculture was also inversely related to change in road density, such that areas of high road development have less agricultural area, and this seems feasible if we consider roads to be a cause of development. In locations where roads have already developed (i.e. low changes in road density), we see high levels of agriculture (behind the arc of deforestation) but where they are developing, agriculture has yet to fully expand. Credit for agriculture is positively related to road density development and it is possible that this positive relationship is linked to the amount of capital available generally for development, thus in areas where there is high credit for agriculture there are also capital resources available for road development. The negative relationship between the percentage of land area protected and road development highlights the role that protected areas play in determining road network development, with highly protected areas experiencing less network growth. While we found that official roads, in conjunction with agriculture, have a significant relationship with road development, they explain little variance in observed road growth over the study period. Further the majority of socio-economic factors tested showed no significant relationship with road development. This, in addition to the fact that socio-economic factors are extremely difficult to predict, they, unlike initial density, do not form a suitable basis upon which to model future road development trajectories.

The temporal dynamics of road network development we describe above and simulated (Figure 5.5) are similar to those of economic boom and bust development trajectories that have previously been observed in the Brazilian Amazon. The development of the Amazon has been economically dependent on boom and bust cycles of extractive industries (Godfrey 1990), with that dependence mirrored in many development indices in the Amazon, including life expectancy, literacy and standard of living (Rodrigues *et al.* 2009). It is thought that the boom in development and improvement in living standards occur as a result of people taking advantage of resources that become available in frontier areas where new roads have been laid (Fearnside 2008, Perz *et al.* 2008, Rodrigues *et al.* 2009).The temporal scale of economic boom and bust cycles appears to match the temporal scale of road network development that emerged from our simulations. For example, the rubber boom of the late 18th century lasted approximately 50 years (Godfrey 1990), cacao booms lasted between 20 and 40 years (Clough *et al.* 2009), and a boom in animal skins from Rio Preto lasted 20 years (Macedo & Anderson 1993). By comparison, our simulations suggested that the boom period of road development lasts for an average period of 39 years. Our model, then, appears to have generated realistic temporal trajectories of road network development for the Brazilian Amazon.

On average it was found that it takes 75 years for an area to reach 'maximum' road density. Maximum is a relative term, as some areas may, in reality, ultimately reach marginally higher densities than those predicted by the simulation. Also, the simulation assumes that all areas will eventually reach the maximum density, whereas many places will not for various reasons including areas that may be protected or where resources are depleted quickly with little investment and thus little road development. Our simulation, then, reflects a scenario in which development continues unhindered until the maximum road density is attained. Further, the rate at which the road network is developed is likely to be highly variable because of differences in geographic environments and economic investment. Our simulation is based on empirical growth rates and as such implicitly takes into account such variable conditions. This is reflected in the range of length of time for road networks to develop (58-103 years).

The Amazon is an ecosystem that provides a multitude of ecosystem services that are globally important (Foley at al. 2007), but is undergoing rapid transformation as forests are progressively cleared for agriculture. As such it is unsurprising that many models have been developed to predict future land uses of the Amazon. However, a key aspect of these models that has been largely ignored is the rates and patterns of road network development (Barlow *et al.* 2011). We have used a combination of spatial and temporal data to explore this issue, revealing the remarkable rate at which road networks are expanding and modelling the temporal trajectory of road networks. We anticipate these analyses will form a base for generating integrated land use models that incorporate the economic development of the region, with a view to gaining a more comprehensive understanding of how this globally important ecosystem is changing.

Chapter 6: **Large scale spatiotemporal patterns of road development in the Amazon rainforest**

6.1. Abstract

There is burgeoning interest in predicting road development because of the wide ranging important socio-economic and environmental issues that roads present, including the close links between road development, deforestation, and biodiversity loss. This is especially the case in developing nations, which are high in natural resources, where road development is rapid and often not centrally managed. Characterisation of large scale spatiotemporal patterns in road network development has been greatly overlooked to date. Here we characterise the spatiotemporal dynamics of road density across the Brazilian Amazon region. We also assess the relative contributions of local versus neighbourhood effects for temporal changes in road density at regional scales. To achieve this we use a combination of statistical analyses and model-data fusion techniques inspired by studies of the spatiotemporal dynamics of populations in ecology and epidemiology. Our results imply that the emergent process of development can be approximated by a simple logistic local growth and spatial dispersal process. We infer the current rates and dominant direction of development and estimate it would take an average of 60 years for roads to spread across the Amazon and achieve maximal densities.

6.2. Introduction

Roads are an important and necessary part of everyday life for most people, forming the basis of the overland transportation network (along with railways) in nearly all countries. Road development influences a wide range of phenomena, from human society, business and economies, to the natural environment (Forman *et al.* 2003). In regional development, roads are often perceived as the initial stage of development, especially in tropical areas where they open access to remote areas for colonisation, agricultural development, and resource extraction (Laurance *et al.* 2001, Arima *et al.* 2005, Perz *et al.* 2007, Caldas *et al.* 2010). Roads further facilitate development by providing market access for rural producers, integrating economic sectors and reducing the cost of spatial mobility (Perz *et al.* 2007). In contrast to the clear positive influence road development often has on human enterprises, it has many and varied effects on the environment. Examples include, but are not limited to: fragmenting habitats and altering their structure, increasing the ratio of edge to non-edge habitats and by extension edge effects, altering animal behaviour, movement patterns and habitat use, altering abiotic conditions, and introducing pollutants (for more details on road effects see: Forman et al 2003, Forman 1998, Forman & Alexander 1998, Spellerberg 2002, Coffin 2007). The majority of these environmental impacts from road development are negative or have negative consequences for existing biota, with one estimate suggesting that impacts on fauna are five times more likely to be negative than positive (Fahrig & Rytwinski 2009). Banning road development in important and delicate tropical areas has been suggested as a way to prevent these negative impacts (Laurance *et al.* 2009). However this is highly unfeasible given the socio-economic benefits and development potential that roads bring (Maki *et al.* 2001).

Roads are constructed for many and varied purposes, endowing them with a wide range of political, topological and morphological differences. One such distinction is the difference between official and unofficial roads. Official roads are built either by the government or with government permission, whereas unofficial roads are built with no planning permission obtained from the state, often by non-state actors, such as miners or loggers (Perz *et al.* 2007, Brandão & Souza 2006). In the Brazilian Amazon the majority of roads built are unofficial and as such there is a distinct lack of spatial information on the location and extent of these roads (Brandão & Souza 2006). This presents a problem for policy makers and conservationists who need spatial information on current and future roads in order to assess potential impacts and make informed decisions.

Development infrastructure, including road and rail networks, is often incorporated into land use change models. These project future land conversions with a view to quantify future changes in carbon flux, climate change and biodiversity. At a global scale two key models are IDRISI Land Change Modeller (Clark labs 2007) and GLOBIO (Alkemade *et al.* 2009). GLOBIO makes predictions of biodiversity change based on five drivers 1) land use, 2) nitrogen deposition, 3) infrastructure, 4) fragmentation, and 5) climate change. Because there is ‘considerable uncertainty’ in predicting growth of infrastructure, future infrastructure development in GLOBIO is based on potential scenarios of road development, for example reducing road growth by 50% or increasing growth by 200%, by 2050 (Alkemade *et al.* 2009). The IDRISI Land Change Modeller on the other hand allows users to incorporate planned road developments into the model. The model also has a road builder module that utilises neural growth dynamics to predict the location of future roads (Jiang 2007). Operating at more local scales, for example in the Amazon, many models predict future land uses by incorporating the location of roads (for examples see; Messina & Walsh 2001, Deadman *et al.* 2004, Walsh *et al.* 2008, Mann *et al.* 2010). However the vast majority of Land use change models do not attempt to predict the future development of the road network and treat the dynamic development process as static (Rosa *et al.* 2014).

There are many factors that influence road development. At large scales, economics, policy, technology, demographic and cultural factors, influence the rate, location and extent of road development (Geist & Lambin 2002, Montagnini & Jordan 2005). The economic climate has a clear influence as it can determine how much capital is available for investment in infrastructure. Government policies greatly influence investment in roads. Governments are likely to provide subsidies to road builders to an area or indeed directly invest in the network by building federal roads to stimulate economic growth. This process was exemplified by the drive to colonise new areas in the Brazilian Amazon in the 1970's (Carvalho *et al.* 2002). Technological advancements influence the cost effectiveness of investments in the road network. Demographics play a role because as a population increases a better infrastructure is required (Glover & Simon 1975). Cultural factors include attitudes, values and beliefs towards roads that might influence development; for instance, many people would be against a road being built through a nature reserve (e.g. Dobson *et al.* 2010). At smaller scales, road alignment (the location of the road in relation to the surroundings) is dependent on a range of factors that can either constrain or facilitate the laying of the road and include topography, existing land use, hydrological features and ground conditions (Koorey 2009).

As a result of the myriad of important socio-economic and environmental issues that roads present, including close links between road development, deforestation, and biodiversity loss, there is an interest in predicting road development (Arima et al 2005). This is especially the case in developing nations, which are high in natural resources, where road development is rapid, and often not centrally managed. Road development predictions are used to aid environmental impact assessments, producing likely scenarios of environmental change and associated impact estimations, such as predicting future biodiversity levels (Alkemade *et al.* 2009, Soares-Filho *et al.* 2006). As a result of the interactions between the transport system and the economic system, economics often play a role in predicting road development patterns. However interactions are complex and subject to time lags, stochastic decisions and feedbacks (Ralston & Barber 1982). This resulted in early models that were descriptive rather than analytical or predictive, that were complicated, and often had poor predictability of road development (e.g. Taaffe *et al.* 1963 and Rene 1964).

More recent models that predict future road networks are often at the spatial and temporal resolution of individual roads. For example, the modelling platform DINAMICA (Soares-Filho *et al.* 2006) incorporates government planned changes to the road network into a model and a secondary road generator is used to predict road development based on land attractiveness. Arima *et al.* (2008) use a similar approach and attempt to predict the location of roads in the Amazon using least cost paths from logging destinations to the main network. IDRISI's road builder module utilises neural growth dynamics to predict future roads (Jiang 2007). These models allow the location of specific roads to be determined; however the validity of these predictions is not well established (Rosa *et al.* 2014). Further, given the current predictive modelling as applied to road development (based on individual roads), the characterisation of larger scale spatiotemporal patterns has been overlooked. Without these larger scale dynamics we miss out on the larger scale rates of change of road development, the general patterns, and what the implications of those patterns might be. Models characterising broad scale dynamics can be easier to generalise and incorporate into other analyses, such as being incorporated into larger models (for example, integrated assessment models) or being applied to other regions. It has been found that approximately two thirds of papers predicting land use change in the Amazon region use roads as a predictor of future land use (Rosa *et al.* 2014). However, the majority of these land use change models treat road development as a static phenomena (given the rate at which roads change, this is not realistic). Thus models that can easily characterise road development could play a vital role in future land use modelling.

Given the complex dynamics and interactions of road development with economics, policy, technology, demographic and cultural factors, it could be asked whether these complicated interactions and driving forces have emergent properties that can be used to predict road development and be described using simple models? Recently Ahmed *et al.* (2013) showed that the dynamics of road density through time in regions of the Amazon can be characterised as a logistic growth curve, where road density initially grows through time at an approximately exponential rate but slows as road density approaches a maximum for that region, suggesting that the road development within a region does have general aspects to its behaviour. This raises further questions about other general characteristics of large scale road dynamics. If roads tend to show logistic dynamics within regions, how then are the regional dynamics related across the Amazon region? *A priori* we would expect the spatiotemporal dynamics to reflect the general direction of development in the Amazon region from the densely populated regions in the east towards the Amazon rainforest in the west. However to date no-one has characterised these dynamics formally.

Here we characterise the dynamics of road density across the Brazilian Amazon, quantifying the spatiotemporal dynamics in terms of the dominant direction and rate of road development. Further we assess the relative contributions of intrinsic versus neighbourhood effects for temporal changes in road density at regional scales. To achieve this we use a combination of statistical analyses and model-data fusion techniques inspired by studies of the spatiotemporal dynamics of populations in ecology and epidemiology (Hilborn & Mangel 1997, Ferrari *et al.* 2008, Haynes *et al.* 2009). Coupling the effects of regionally local dynamics with spatial dispersal using simple mathematical functions to predict and characterise the spatiotemporal dynamics of populations has proven to be an extremely useful approach in ecological research (Murray 2002, Sherratt & Smith 2008). Using simple phenomenological functional forms (such as the logistic equation for population growth) enables detailed investigation into the possible emergent spatiotemporal dynamics arising from simple underlying principles about population birth, death and dispersal processes without initially becoming too distracted by the myriad details underlying these processes. Further details can then be subsequently incorporated when there is objective justification that their inclusion leads to sufficient improvements in model predictive accuracy. Statistical analyses of the existing empirical data of the system(s) of interest usually accompany model building to aid in the identification of appropriate functional forms. Parameter values representing vital rates, such as birth, death and dispersal rates are typically either derived from other studies, expert knowledge or estimated using formal inference methods. In general we follow similar principles here: preceding modelling with statistical analyses to characterise the existing spatiotemporal dynamics, and then using simple spatiotemporal population models with simple functional forms, parameterised using parameter inference, to enable efficient characterisation of the emergent spatiotemporal dynamics in terms of the rate and direction of road development at regional scales.

6.3. Methods

6.3.1. Data and general pattern analyses

Data on road density (km roads/km^2 land area) for 443 municipalities of the Brazilian Amazon in 2004 and 2007, were the spatial units for analysis (Figure 6.1). These data were derived using manual digitisation methods on Landsat TM imagery of the Brazilian Amazon (see Ahmed *et al.* 2013, for details). There are many automated approaches to digitising road networks (see Mena 2003, Brandão & Souza 2006, Li & Briggs 2009, Movaghati *et al.* 2010), but these are typically less accurate than manual digitisation (Li & Briggs 2009) in which images are observed and visible roads are traced individually and by hand. All visible roads are recorded using this method and any additional information known about the roads, for example whether they are primary or secondary roads, are added to the data set.

Before choosing models we carried out preliminary analyses of the data to look for patterns that would help determine, in conjunction with a priori knowledge of processes in the region, the most appropriate models. Based on previous work carried out at the municipality scale we know that roads appear to follow a density dependent development pattern in which the rates of growth per capita (km) are negatively related to road density (Ahmed *et al.* 2013). In low density municipios there is little additional road development, probably as a result of low investment and high access costs, in mid-road-density municipios road development is at its highest rates, and in high-road-density municipios growth slows, possibly due to a loss of economic impetus to expand. These analyses showed that road development is affected by the existing road density of a given municipality, but here we extended on this concept and carried out a General Linear Model Analysis (GLM (Crawley 2008))to determine if road density in neighbouring municipalities had an influence on road density growth in a given municipality.

Moran's I was calculated for the whole study area to investigate road density anisotropy (spatial correlations) among municipalities (Legendre & Legendre 1998). Further, directional anisotropy was used to determine if the development of road networks was moving in any particular direction. We would expect that development moves from south to north and from west to east, from the arc of deforestation towards 'new' forest rich areas with resources. Directional anisotropy was carried out by projecting municipality centroids (and corresponding densities) along multiple directional planes, separated by 10 degree increments. Our assumption was that the direction with the spatial auto-correlation extending the shortest distance is the direction that the growth is extending to. Conversely the direction in which directional anisotropy extends the furthest is *not* the direction in which road development is moving. We carried out directional anisotropy analysis at two 'scales'; Amazon wide and quadrant, where we divided the Amazon into four quadrants and analysed each separately to see if the directional patterns observed at an Amazon wide scale held at smaller scales.

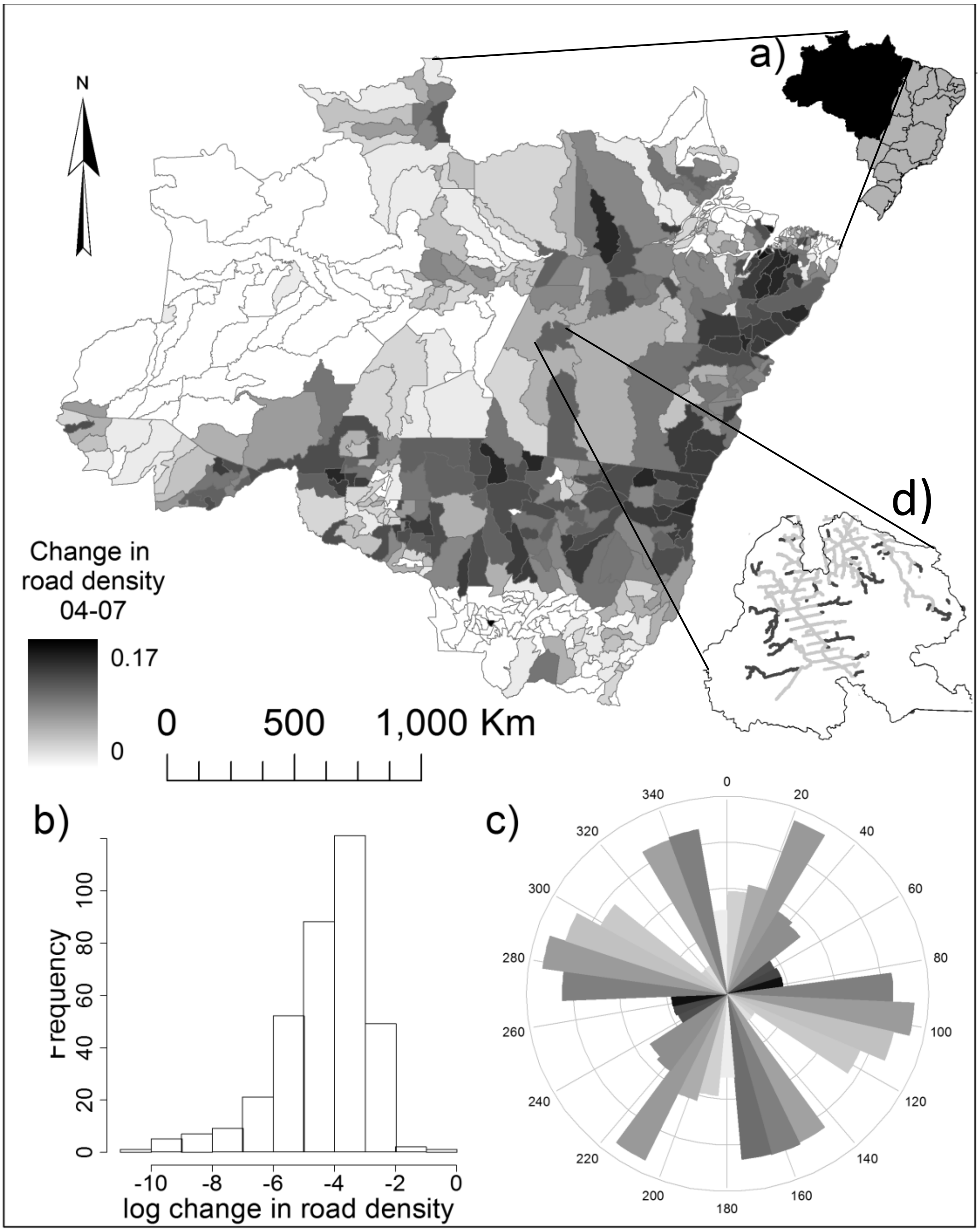


Figure 6.1. The spatial distribution of the change in road density between 2004 and 2007 in the Amazon legal, divided by municipio. Change in road density is concentrated along the arc-of-deforestation a) Location of study site within Brazil, with Brazilian state lines shown. b) Histogram of log-change in road density between 2004 and 2007. c) Directional anisotropy radial plot, displaying the extent to which statistically significant spatial correlation in change in road density extends in different directions. 'Long' bands indicate correlation extends to greater distances, indicating that road development is moving at a perpendicular angle (different grey tones are a visual aid to enable differentiation between bands). d) Example of road development over a three year period between 2004 (light grey) and 2007 (dark grey).

6.3.2. Models

We consider a set of 16 simple models for the spatio-temporal growth of road density and assess how well each of these explain the observed data given the inferred probability distributions for its parameters. For all models we predict the growth in road density for each location between the years 2004 and 2007. The median percentage increase at the municipio level for these sites over that time window is 10% (90% intervals: 0% and 60.89%). This relatively small magnitude of change implies that we might obtain a good approximation of the increase in road density over that time period simply from predicting the rate of change in 2004 using an ordinary differential equation formulation and then assuming a constant rate of change over the time window. We experimented with solving our models using this assumption and solving them using smaller time steps; both approaches obtained qualitatively identical results. Therefore, for convenience, we chose to solve our ordinary differential equation models using the basic forward Euler method with a one year time step (thus, three time steps per solution). This made it natural to extend the implementation to study the dynamics for longer time-period model simulations.

We applied the 16 models to three sets of data each, the first data set is road density divided by municipality. The second and third data sets are road density divided by equal area grids of $100km^2$ and $50km^2$.These were adopted to investigate the effects of the spatial resolution (scale) and the effects of the irregularly sized municipalities on our results.

Travelling wave model

We studied a model in which the spatiotemporal dynamics of secondary road growth in the Brazilian Amazon is described as a travelling wave, with zero or low road density in front of the wave, the highest road density behind the wave and the moving "wave front" capturing the change in road density through time and space. We assume a simple travelling wave form: a logistic equation in shape and moving in one direction only at a constant speed. Our statistical spatial analysis of the road data detailed below implies that there are likely to be multiple wave fronts travelling in different directions in the Brazilian Amazon. For our analysis we divided the Brazilian Amazon into the same 4 quadrants as used in the statistical spatial analysis and assessed evidence for separate travelling wave phenomena in each, assuming a unidirectional travelling wave in each quadrant. We also used the wave model on the Amazon as a whole (without dividing it into quadrants). Our travelling wave model is

$$\rho_m(t) = \frac{K_w}{1+\exp\{-R_q[\gamma_q(\alpha_q,\omega_q)+\tau c_q]\}}, \quad (1)$$

where ρ_m is the road density (km km^{-2}) in location (municipio or grid cell)m at time t (years), K_w is the maximum road density behind the wave front (km km^{-2}), R_q scales the rate of change of road density with space (km km^{-2}y^{-1}) for quadrant q, $\gamma_q(\alpha_q, \omega_q)$ is the distance from the midpoint of the location to the midpoint of the travelling wave with coordinates (α_q) when projected along the angle of movement of the travelling wave relative to north (ω_q,), τ is the number of years that have elapsed since 2004 and c_q is the speed of travel of the travelling wave (km y^{-1}). Note that we assumed K_w to be the same for each quadrant when fitting this model to the different quadrants separately.

No neighbourhood effects models

For the first two models not incorporating a moving wave we assume that the change density of roads in a municipio is solely a function of the road density in that municipio previously in time. In other words, we assume no neighbourhood effects. The first model we considered was that previously implied by the study of (Ahmed *et al.* 2013) in which the change in road density through time is modelled as a logistic growth process. To use a numerical solution method that is identical to the subsequent models we simulate this process using the ordinary differential equation

$$\frac{d\rho_m}{dt} = \rho_m r\left(1 - \frac{\rho_m}{K}\right) = Logistic \qquad (2)$$

Where K is the maximum road density per location and r is the maximum rate of growth of road density through time (km km^{-2} y^{-1}). Note that for this formulation we assume K and r to be the same for all locations even though we expect both to vary between locations, however our focus with this study was to investigate the predictive performance of the simplest models.

The second model we consider predicts the change in road density through time as an exponential growth process

$$\frac{d\rho_m}{dt} = \rho_m r = Exponential \quad (3)$$

where the parameters and assumptions are as defined above. This model was investigated simply to contrast a model that assumed density dependence in the growth of road density (equation 2), with one that did not.

Neighbourhood modules

Given a GLM indicated that neighbouring municipio road density affected change in road density (details below), for the next 12 models we predict the growth of road density in each location as a combination of local growth and neighbourhood effects. We model these as independent processes thus all 12 models have the general form

$$\frac{d\rho_m}{dt} = Local\ growth(Logistic) + Neighbourhood\ effects \qquad (4)$$

All of our preliminary analyses supported the use of the logistic model (equation (2)) rather than the exponential model (equation (3)) as the local growth process and so we used equation (2) to model local growth in equation (4) for all of our nearest neighbourhood models. The models detailed here therefore only differ in their neighbourhood effects component.

We formulated the different neighbourhood effects components by reasoning that neighbourhood influences will probably vary as a function of the relative difference in road density between neighbouring locations. For these models the neighbours of any location are all those that share borders with it. We hypothesise that the larger the difference in road density between neighbouring locations, the larger the pressure will be on the neighbour with the lower road density to increase in road density. We also hypothesised that this effect may not be reciprocal; so a location with high road density neighbouring one with low road density may not feel any neighbourhood effects but will be dominated by local growth processes. We considered four different neighbourhood effects formulations (referred to as NEm1-NEm4)

$Neighbourhoodeffects_m = D_1 \sum_{j=1..n} F(\max(\rho_j - \rho_m - \tau, 0))$ (5, NEm1)
$Neighbourhoodeffects_m = D_2 \sum_{j=1..n} F(\frac{\max(\rho_j - \rho_m - \tau, 0)}{E_{mj}})$ (6, NEm2)
$Neighbourhoodeffects_m = D_3 \sum_{j=1..n} \rho_j F(\max(\rho_j - \rho_m - \tau, 0))$ (7, NEm3)
$Neighbourhoodeffects_m = D_4 \sum_{j=1..n} \rho_j F(\frac{\max(\rho_j - \rho_m - \tau, 0)}{E_{mj}})$(8, NEm4)

where D scales the magnitude of the neighbourhood effect (units differ depending on formulation), j is the identity of one of the n neighbours to location m, E_{mj} is the Euclidean distance (km) between the centroids of locations m and j, $\tau$is the threshold difference in road density between a neighbour and the focal location, below which the focal location experiences no neighbourhood effects, and F is one of three different functional transformations. This threshold can be positive, in which a neighbour must have a larger density than the focal cell to exert a positive effect, or negative or zero.

The first of the four formulations above (equation (5)) predicts the size of the neighbourhood effect as a function of the sum of positive differences in road density between a location and its neighbours. The second (equation (6)) assumes that there is an additional effect based on the relative distance between neighbours, approximated as the Euclidean distance between their centroids. The third (equation (7)) assumes that the size of the neighbourhood effect is a function of the difference in road density between a location and that of its neighbour and of the absolute road density of that neighbour. The fourth (equation (8)) assumes this and an additional effect from the Euclidean distance between neighbouring sites.

Dispersal type

We considered three different functions, F, for use in equations 5-8 above, representing different hypotheses about the relationship between the difference in neighbourhood density and the size of the neighbourhood effect. These are illustrated in Figure 5.3. The first is trivially "no transformation", of the form $F(y) = y$. The second is an exponential function of the form $F(y) = \exp(y)$. The third is a saturating function of the form $F(y) = y/(\varphi + y)$ where φ is a half saturation constant. Representations of these different functional forms, drawn using parameter probability distributions inferred from the empirical data and reported in the results, are given in Figure 6.2. These simple functional forms were selected as simple first representations of density dependent neighbourhood effects.

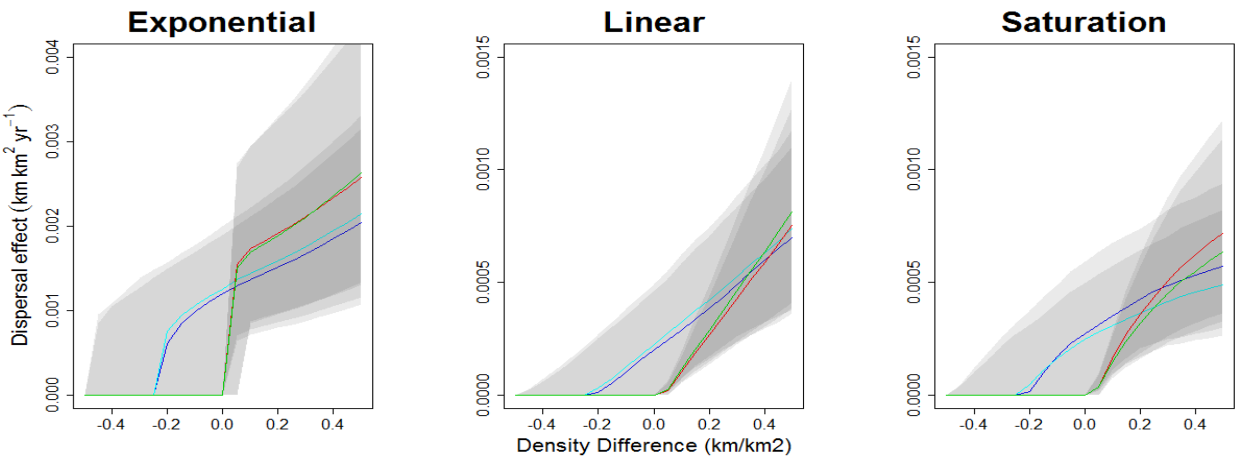


Figure 6.2. Mean dispersal effects graphs for 4 neighbour models (equations 4-7 with Exponential, Linear or Saturating functional forms as defined in section 2.3) with 95% confidence intervals (grey bands). Note, dispersal effect magnitude change between dispersal types. Lines; red=NEm1, green= NEm2, dark blue= NEm3, light blue= NEm4.

6.3.3. Maximum likelihood parameter estimation

We inferred the probability distributions of the parameters to the above models given the data using maximum likelihood parameter estimation. We expect that most of the variance in the observed change in road density derives from inherent stochasticity in the road development process rather than any error in taking the measurements of road density which we are confident are accurate (Brandão & Souza 2006, Ahmed *et al.* 2013). We therefore expect larger variation amongst sites that showed relatively large changes in road density over time due to the expected multiplicative nature of the road growth process. Preliminary analysis of the data also indicated lognormal variation in the change in road density across all sites. Therefore for all models, except the travelling wave model (for reasons detailed below) we estimate the likelihood of the data given the predictions of a parameterised model using

$$\log\left(L(Model(\boldsymbol{\theta})|\boldsymbol{X})\right) = \sum_{j=1..J} \log(\mathrm{P}(\log\left(\rho_j(2007) - \rho_j(2004)\right) \approx \aleph(\log\left(\Delta\rho_{j,pred}\right), \sigma^2)) \quad (9)$$

where L is the likelihood of the model with vector of parameters $\boldsymbol{\theta}$ given a vector of data, $\boldsymbol{X}$ of length J, P is the probability that the log of the observed change in road density between 2004 and 2007 is drawn from a normal distribution, $\aleph$, with mean centred on the log of the predicted change in road density over that time period, $\Delta\rho_{j,pred}$, and variance σ^2 which is estimated alongside the other model parameters. Hence, we infer the most likely parameters by assessing the difference between the predicted and observed log of the difference in road density between 2007 and 2004 for every site. We excluded sites that exhibited no change in road density over this time window. Note that this is effectively making a “space for time” assumption: differences in road density changes over the time window at different points in space are assumed to occur solely due to differences in their initial conditions. We discuss the implications of this assumption in the *Discussion*. For the travelling wave model we used the 2004 and 2007 data to infer the most likely position and parameters of a unidirectional travelling wave in each of the 4 quadrants.

$$\log\big(L(Model(\boldsymbol{\theta})|\boldsymbol{X})\big) = \sum_{\tau=0,t=3}\sum_{j=1..J}\log(\mathrm{P}(\log\left(\rho_j(2004+\varepsilon)\right) \approx \aleph(\log\big(\rho_{j,pred}\big), \sigma^2))$$
(10)

Where $\varepsilon = 0$ corresponds to the 2004 road density data and $\varepsilon = 3$ corresponds to the 2007 road density data. The only thing that changes in equation (10) between those two time points is the ε parameter, which enables us to infer the wave speed parameter m_q. Inference of the travelling wave model was performed using the raw data rather than the change in road density because it enabled us to infer the speed and direction of the travelling wave.

For all models we used Markov Chain Monte Carlo sampling with the Metropolis-Hastings algorithm (Gilks *et al.* 1996) to perform the parameter estimation which we implemented using the Filzbach libraries (http://research.microsoft.com/en-us/um/cambridge/groups/science/tools/filzbach/filzbach.htm) within a Microsoft Visual Studio C# solution. We generated fake data sets using the models above and selected parameters to confirm that we could recapture the parameters using these algorithms. All Markov Chains were 100,000 iterations in length after a 10,000 iteration burn in period, which we confirmed to be sufficient for parameters to converge to their posterior distributions. Markov Chains were sampled every 100 iterations to remove autocorrelation and so the posterior parameter estimates are calculated from 10,000 Markov Chain samples.

We used 10-fold cross validation to assess the sensitivity of our inferred parameter probability distribution to different subsets of the data and to assess the performance of the fitted models against data not used in training (to prevent over-fitting). We randomly assigned each datum to one of 10 folds. We then removed all data assigned to one fold (the evaluation data) and performed maximum likelihood parameter estimation on the remaining data (the training data). Model performance was assessed using the withheld fold of data. Repeating this procedure for each fold of data generated 10 sets of maximum likelihood parameter estimates and 10 sets of model performance estimates.

6.3.4. Data analysis

In addition to analysing the posterior parameter estimates for the different models we assessed the goodness of fit of the model to the data using five different metrics: the Deviance Information Criterion (DIC, (Gelman *et al.* 2004)), the correlation between the model predictions and the evaluation data (CC), the coefficient of determination between the model predictions and the evaluation data (CD), and the mean log likelihood of the training (TL) and evaluation data (EL).

6.4. Results

6.4.1. General patterns

The change in road density between 2004 and 2007 was approximately log-normally distributed (Figure 6.1b). Spatially the distribution of the change was concentrated along the arc-of-deforestation (Figure 6.1a). A GLM indicated that the growth in road density is significantly positively related to the initial density of a municipio ($p<0.05$, slope=0.11) and also indicated a significant interaction effect between the initial density of a given municipio and its average neighbourhood road density on the growth of road density ($p<0.05$, d.f.=470, slope=-0.36). However, average neighbour density alone was not significantly related to the change in road density ($p=0.18$).

A Moran's I test indicated that there was significant autocorrelation in road density change between municipalities that were up to 434 km apart (Figure 6.3). The analysis of spatial anisotropy in the change in road density between 2004 and 2007 implies a number of different directions of the development of the road network across the whole Amazon with correlation with distance dropping away most sharply in both north westerly and north easterly directions (between 310 and 330 degrees and between 240 and 270 degrees in Figure 6.1c). When the same analysis is performed on the data divided into four separate quadrants then a clearer directionality to road development is apparent (Figure 6.4) and highlights contrasting patterns of directional road development that is in general directed towards the centre of the Amazon, along the arc of deforestation, although a dominant direction is not apparent in the south-west quadrant (Figure 6.4d).

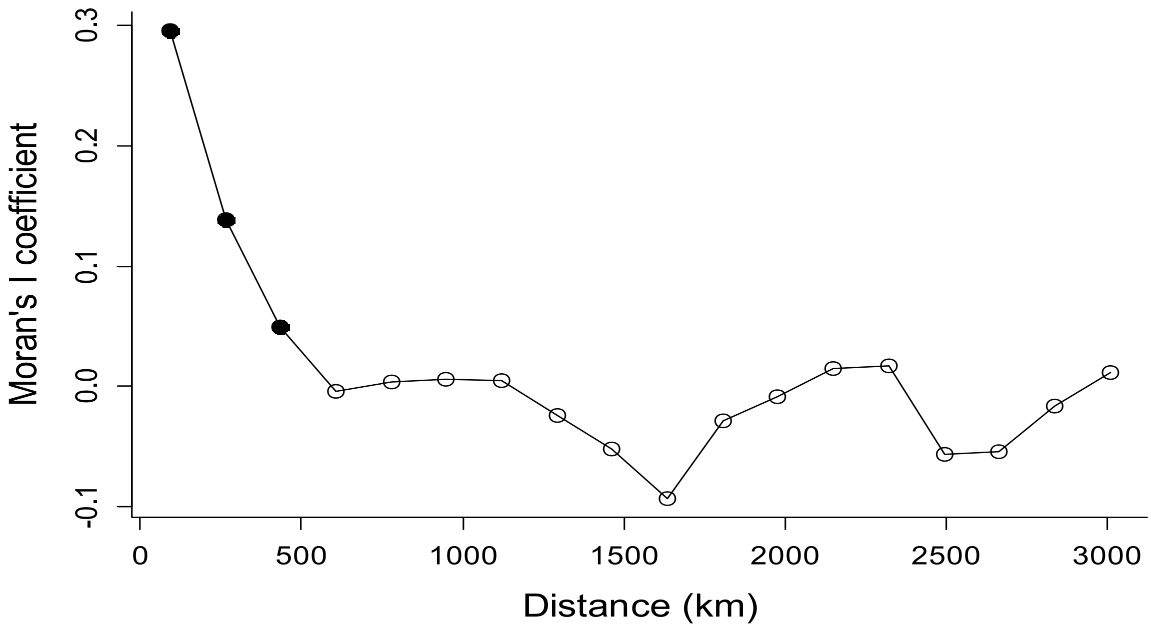


Figure 6.3. Moran's I coefficient was calculated for the whole study area to investigate road density anisotropy (spatial correlations) among municipalities. Filled points indicate significant correlation of road density change between 2004 and 2007 among municipalities up to a given distance (shown on the x axis). Unfilled points represent no significant correlation. There was significant autocorrelation in road density change between municipalities that were up to 434 km apart.

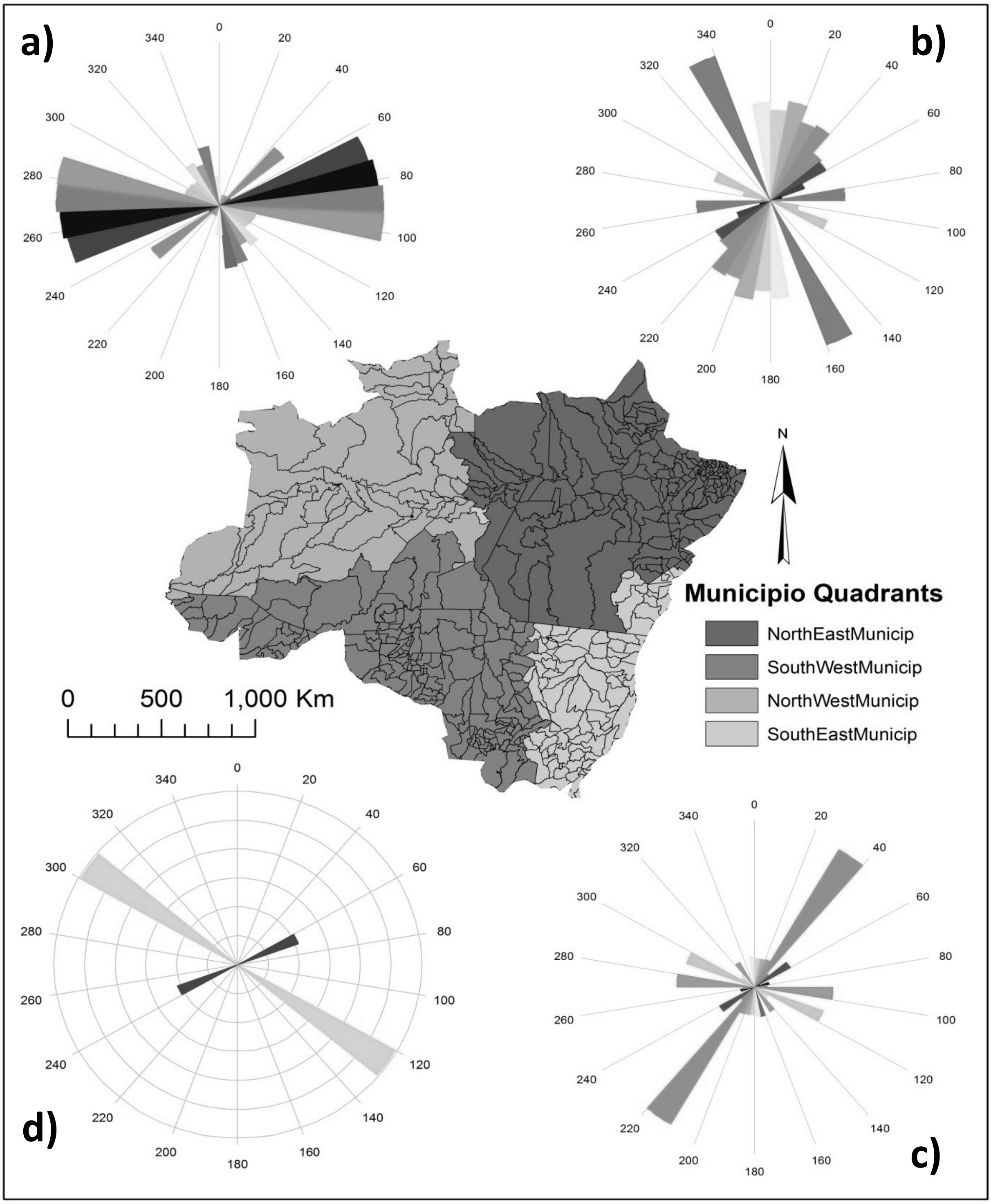


Figure 6.4. Radial plots of directional anisotropy of the Amazon divided by four quadrants. a) north west (NW), b) north east (NE), c) south east (SE), d) south west (SW). The direction in which development is moving is much more pronounced when four regions are considered in contrast to the Amazon as a whole (Figure 6.1).

6.4.2. Models

For the wave models, at all spatial resolutions studied (municipio, grid50 and grid100), the model predictive performance was better when fitted separately to the data divided into quadrants than when the same model was fitted to all of the data together. This is an expected consequence of enabling the model to have more degrees of freedom (effectively more parameters).

Of the three spatial resolutions we studied, the wave model marginally performed worse when fitted to data at the 50km resolution (grid50) than at the 100km or the municipio resolution. For example when the data was divided into quadrants, grid50 had an average correlation coefficient of 0.54 (average 95% confidence intervals of 0.51 & 0.57) for predicting the evaluation road density data compared to and 0.65 (0.62 & 0.69), and 0.59 (0.51 & 0.66) for the municipio and grid100 resolutions respectively. These coefficients indicate significant variance in the relationship between predicted and observed road density and is confirmed by comparisons between predicted and observed road densities, which show a loose correlation between log predicted and log observed road density in 2007 and a bias towards over-prediction of road density at low road densities.

The parameter estimates for the wave models corroborate the findings of the spatial anisotropy analysis; with the dominant direction of road development tending to be perpendicular to the arc of deforestation (Figure 6.5c, for brevity we only present the parameter estimates for the grid100 resolution). The results also indicate the average speed of wave movement as being ~54km yr^{-1}, with the southeast having the highest speed of 114 kmyr-1 and the northwest having the lowest speed of 23 kmyr^{-1}. Given these speeds the wave of road development would take on average 55 years (min=26, max=130 years) to traverse the study region (the region is approximately 3000 km wide along its widest dimension). As with the wave models, the neighbourhood effects models fit at the 50km grid resolution performed worst out of all of the models and thus results for this scale are not presented.

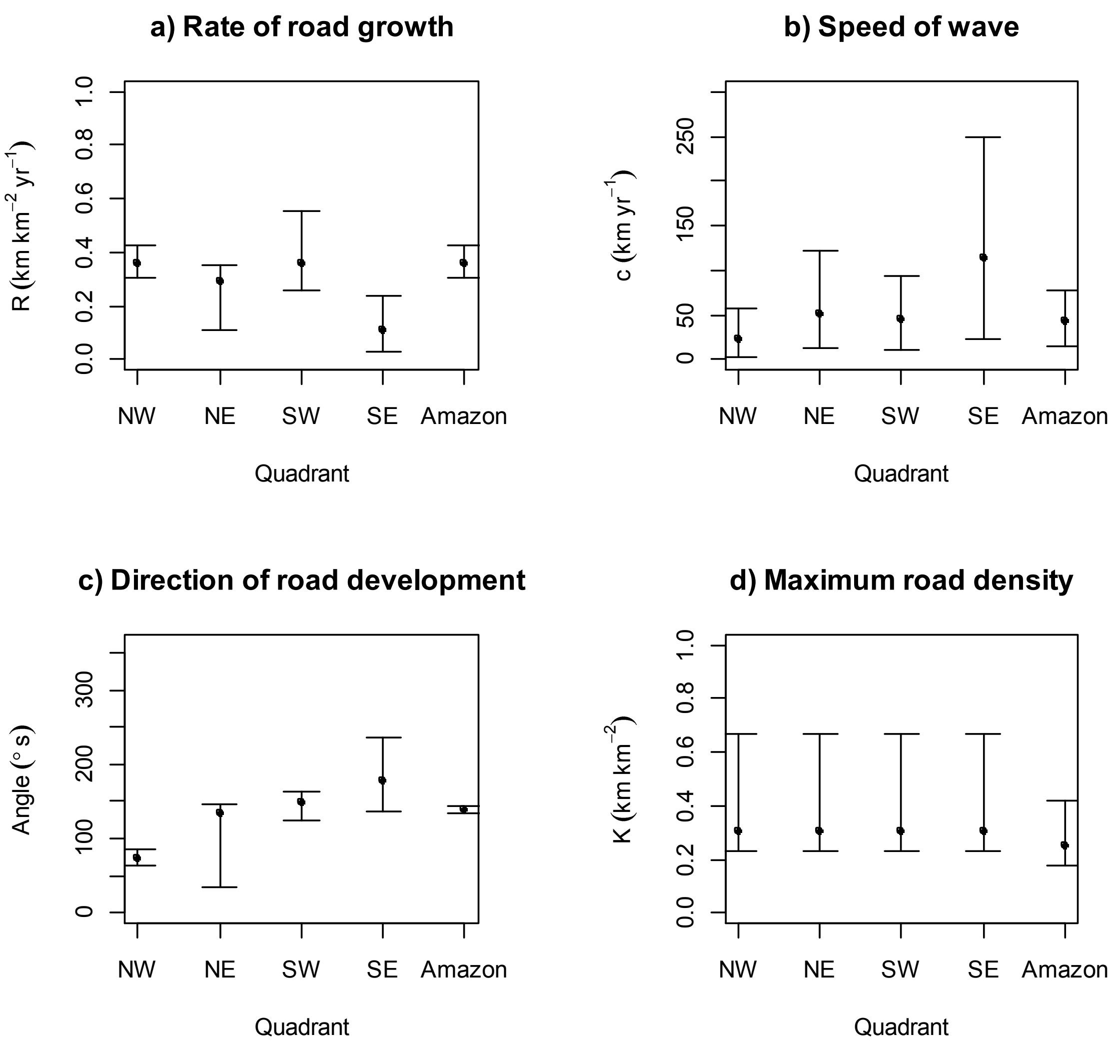


Figure 6.5. Parameter estimation from the Wave model for four parameters, at two scales, 1) Amazon wide and 2) Quadrats (NW, NE, SW, & SE) which correspond to the quadrants displayed in Figure 6.4.Mean parameter estimates and 95% confidence intervals are displayed. R= rate of road density change (km km^{-2}yr^{-1}), c= speed of wave (km yr^{-1}), Angle= angle of the travelling wave relative to north (degrees), K= maximum road density behind wave (km km^{-2}).

There was also little difference between the grid100 and municipio scales, as such we present results at the grid100 scale only. Of the three dispersal types considered (no transformation, exponential and saturation), the exponential dispersal type performed marginally better, thus for brevity, we present results from models with an exponential dispersal type only, allowing a fair comparison between each of the four different neighbourhood effects functions. The model performance metrics clearly indicate the exponential model with no neighbourhood effects (equation 3) to be the worst performing model in terms of its likelihood and ability to predict the evaluation data (Figure 6.6). In contrast, the logistic model with no neighbourhood effects (equation 2) only appears to perform marginally worse than the models with neighbourhood effects, as indicated by a slightly higher DIC (with a mean DIC of 413 compared to mean DICs between 374 and 383) and lower training log likelihoods (TL), with a median TL of -0.8 compared to median TLs of between -0.74 and 0.75. However the correlation coefficient (CC) and the coefficient of determination (CD) indicate a very similar level of predictive accuracy to the models with neighbourhood effects.

The best performing neighbourhood effects models at the 100km resolution predicts the log change of road density in the evaluation datasets with a correlation coefficient of ~0.45 and a coefficient of determination of ~0.2 (Figure 6.6), the latter implying that the model explains ~20% of the variance in the data. There are only minor quantitative differences in the predictive performance of the different neighbourhood effects models (Figure 6.6, see Appendix B for all model results).

The maximum local per capita change in road density, r, tends to be higher for the logistic model with no neighbourhood effects than the neighbourhood effects models (Figure 6.6, we omit further mentioning the exponential model with no neighbourhood effects). This is clearly due to the logistic model inferring higher overall local changes in road density to explain the observed changes compared to the neighbourhood effects models. Similar maximum road densities are inferred for the logistic model with no neighbourhood effects and all of the neighbourhood effects models (Figure 6.6).

Although some of the different neighbourhood effects models appear to show contrasting inferred parameter values, these should be interpreted in the context of the complete functional forms illustrated in Figure 6.2. This illustrates that the different functional forms in the exponential models predict similar magnitudes of neighbourhood effects with the most notable difference being the inference of a positive threshold at which neighbourhood effects occur, τ in Figure 6.6, for the neighbourhood effects models that lack the additional multiplication factor of the local road density (equations 4 and 5, red and green lines in Figure 6.2). However we note that these differences in functional forms only have minor effects on the predictive performance of the model (Figure 6.6). These results indicate that the larger the net difference between two neighbours, the larger the neighbourhood effect is on the neighbour with the lower road density.

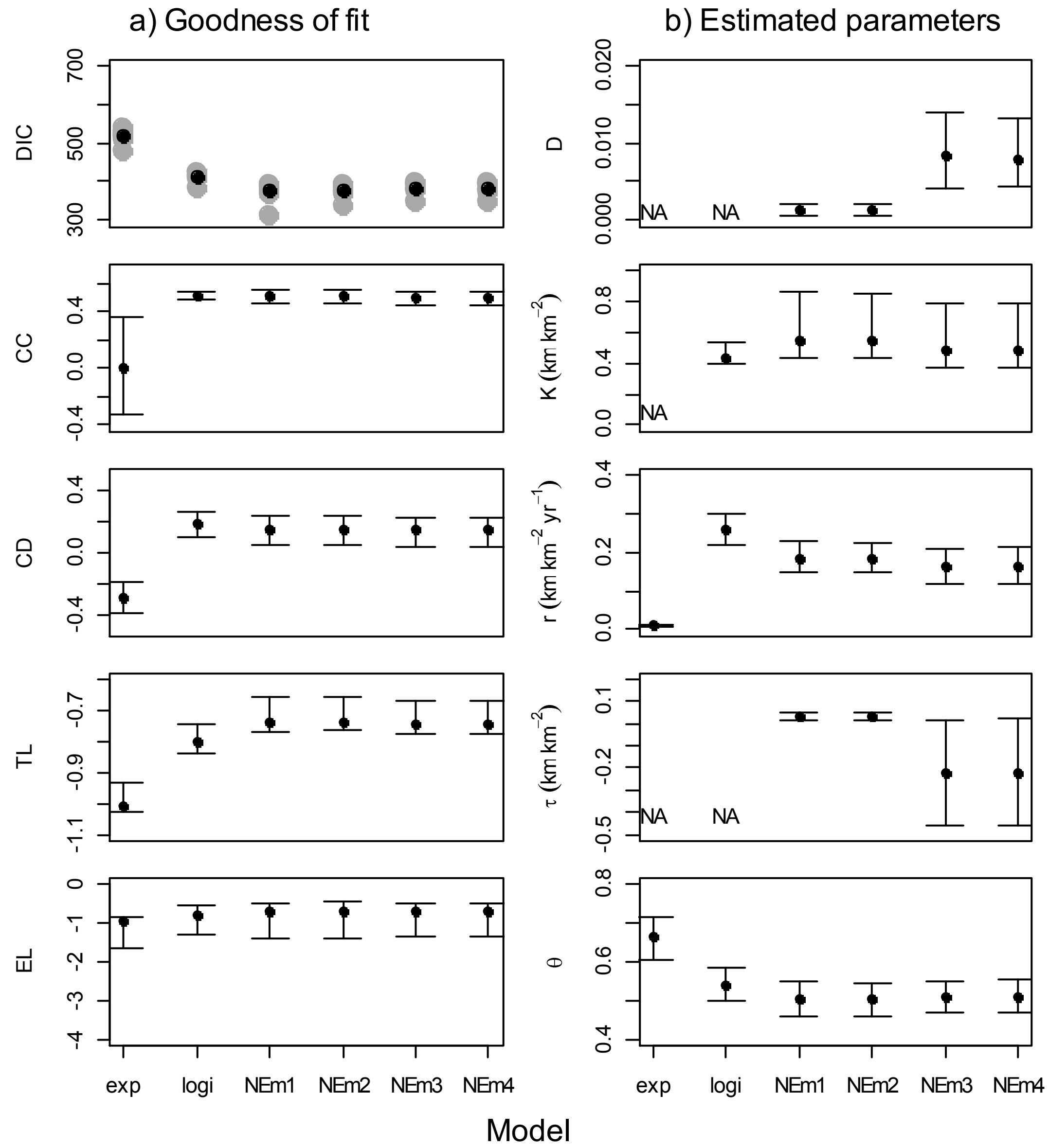


Figure 6.6. a) Goodness of fit measures for logistic (logi), exponential (exp) and 4 neighbourhood effects models (equations 4-7 with Exponential functional forms as defined in section 2.3). Mean parameter values and 95% confidence intervals are displayed for all goodness of fit measures except DIC, for which mean DIC (dark circle) and 10 DICs (grey circles) from each of the 10 fold parameter estimations are displayed. The exponential model performs worst for all measures followed by the logistic model lacking neighbour effects. Model NEm1 performs slightly better than the other three neighbourhood effects models. DIC=Deviance information criterion, CC= coefficient of correlation, CD= coefficient of determination, TL= training likelihood, EL= evaluation likelihood. b) Estimated parameters for logistic, exponential and four neighbourhood effects models (equations 4-7 with Exponential functional forms as defined in section 2.3) at the grid100 scale. Mean parameter values and 95% confidence intervals are displayed. D= magnitude of neighbourhood effect (units differ depending on formulation see section 2.3), K= maximum road density (km km^{-2}), r=maximum road growth rate (km km^{-2} yr^{-1}), τ=road density threshold difference (between neighbours) at which neighbourhood effects become apparent (km km^{-2}), theta = estimated variance in the observations about the model predictions.

Comparing the predictions of road densities using the best wave and neighbour models (marginally the best in the case of the neighbourhood effects models) highlights contrasting abilities to predict log road densities and log changes in road densities. Comparing predicted and observed log road densities for these models indicates a much better predictive performance by the neighbourhood model (Figure 6.7). The NEm1 model predictions were closely correlated with the observed road densities (Spearman's correlation coefficient, $r=0.98$, 95% CI=0.97-0.98, $p=<0.001$, $t=79.0$, $df=283$), while the wave model has a lower, but still significant, correlation ($r=0.54$, 95% CI=0.46-0.62, $p<0.01$, $t=10.9$, $df=283$). The good performance of the neighbourhood effects model is clearly largely because its prediction incorporates the actual road density in 2004 (it only predicts the change over that time window, compared to the wave model that simply predicts the log road density in 2007). Comparing the two models in their ability to predict the log change in road density over the three year period again reveals contrasting performance, while the neighbourhood model still has closer correlation between observed and predicted change in road density ($r=0.65$, 95% CI=0.58-0.71, $p<0.01$, $t=14.5$, $df=283$ compared to $r=0.32$, 95% CI=0.21-0.42, $p<0.01$, $t=5.7$, $df=283$). In the case of predicting change in road density the wave model tends to over-predict the change in road density at low road densities whereas the neighbourhood model is less biased (Figure 6.6b), although it still tends to over-predict log changes in road density at low road densities.

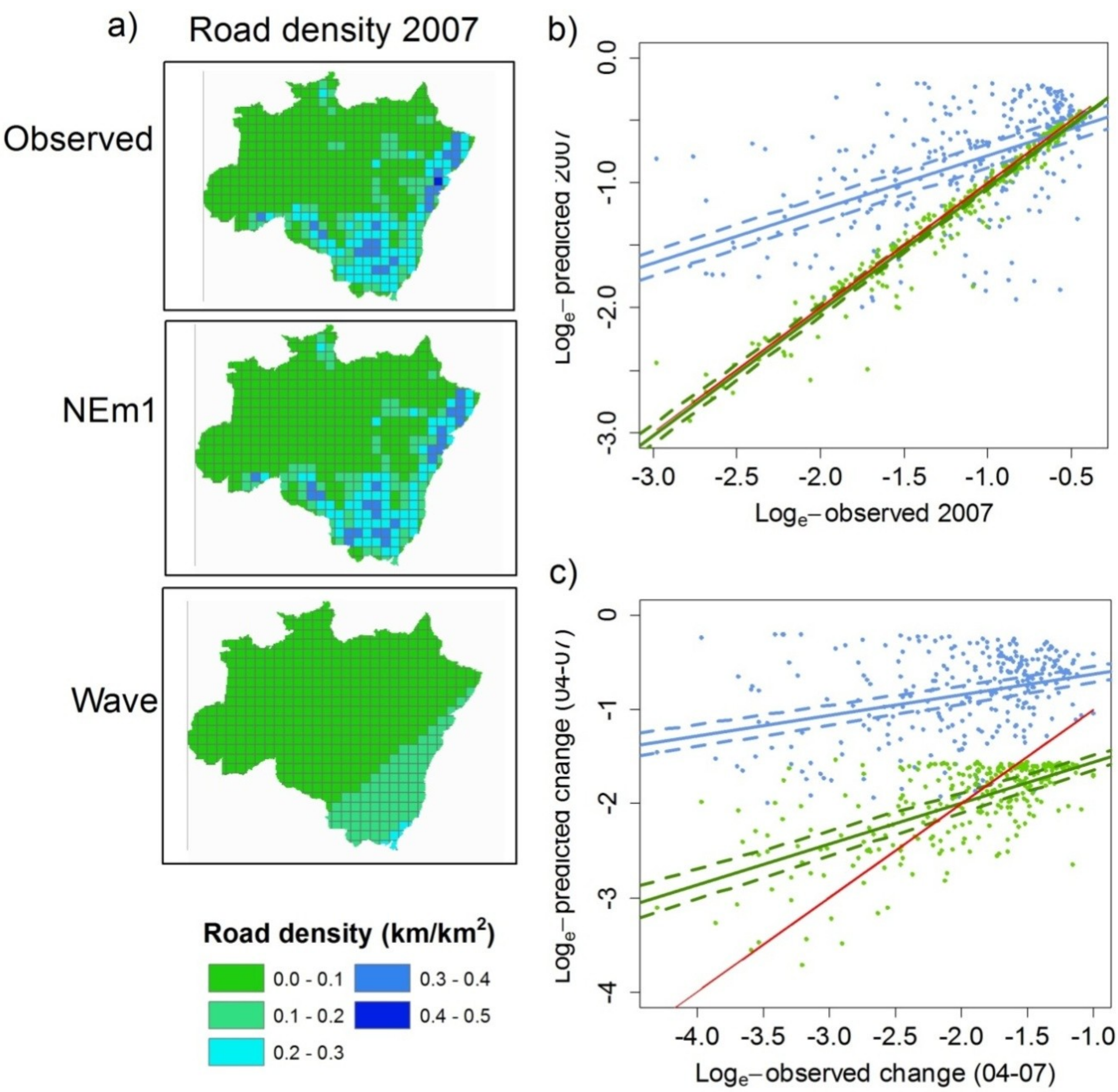


Figure 6.7. a) Observed road density in 2007 and road density in 2007 predicted by NEm1 and Amazon wide wave models. b) Observed versus predicted $\log_e$ road density in 2007 from wave (blue circles) and NEm1 (green circles) based on average median estimates for each location. Correlation lines for each model are displayed (solid lines, wave=blue, NEm1=green). Correlations for upper and lower 95% confidence intervals are also displayed (dashed lines). A 1:1 line is shown for reference (red solid line). NEm1 has better predictions of absolute $\log_e$ road density in 2007. c) Assessment of model predictive accuracy based on observed and predicted $\log_e$ density change between 2004 and 2007 (same colour scheme as in b) is used).

As a final analysis we investigated the projections of the models when extrapolated over longer time frames to assess the predicted rate of spread of road density across the Amazon. These projections should be interpreted with caution for several reasons we address in the discussion. In terms of temporal scale both the neighbourhood effects and the wave models show a degree of similarity with it taking 65 years (based on 90% of grid cells reaching a road density greater than 0.47km/km^2) and ~55 years respectively for road density to reach the maximum predicted density across the entire Amazon (Figure 6.8). Although we find the neighbourhood model predicts maximum road density across the Amazon this is unrealistic because there are many barriers to road development and the model assumes a homogenous environment. We incorporated barriers to road development (rivers and protected areas) on a 100km grid in a simple way (where any grid cell with an area of more than 75% covered by barriers was considered 'protected' i.e. no roads would develop, Figure 6.8). These results indicate vary little influence of the barriers on the rate of road spread and only a few large areas unaffected by road development, although this is most likely due to the coarse spatial resolution we adopted. We then repeated the analysis on a 10km grid by simulating the model as a partial differential equation (details in the legend to Figure 6.8). In this case the large areas of the Amazon remain relatively undeveloped and the rate of advance road density is notably slowed by the barriers to development.

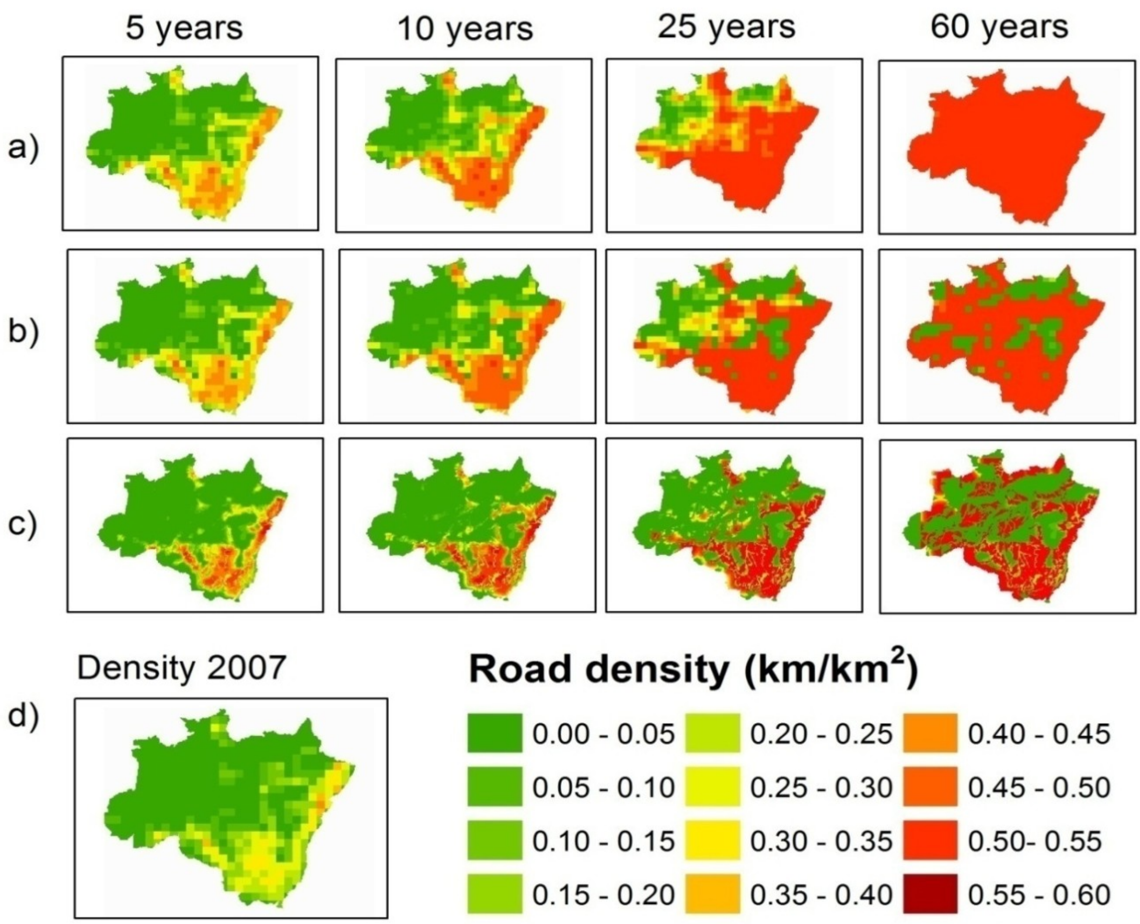


Figure 6.8. a) Future projections of road density modelled on a 100 km grid, based on NEm1 (Equation 5). Estimates suggest that within approximately 60 years the whole Amazon will have a relatively homogenous road density of 0.5 km/km^2. b) Projections of future road density, modelled on a 100km grid, based on NEm1 incorporating barriers to dispersal (road development); rivers and protected areas. A grid cell is considered a barrier when >75% of its area is covered by a barrier (river or protected area). c) Roads modelled on 10km grid; the rate of spread is slowed and more complicated patterns of road development are evident when compared to projections made without real world dispersal barriers and at coarser resolutions. Simulation results in a) and b) are the mean estimates from running simulations for all 10,000 combinations of parameter values for each of the 10 fold model fitting runs. The simulations in c) were made by converting the model to a partial differential equation combining diffusive dispersal and logistic population growth, but where the diffusion rate is determined by the NEm1 model formulation. It was solved using a fully explicit finite-difference method (Smith 1986) with time step of 0.01 years for one draw of parameters from one of the Markov Chains – simply to provide a representative simulation that would illustrate the effects of barriers at finer resolution.

6.5. Discussion

Road network development forms an important part of regional development in Brazil, with the Brazilian government pledging US$500 billion, in 1999, as part of the ‘Avanca Brazil’ (Advance Brazil) initiative. One fifth (21%) of this funding was allocated to infrastructure development including the construction of new roads and paving of existing roads (Carvalho *et al.* 2002). The aims of this development initiative mirror those from the 1960/07s; a desire to integrate the Amazon through colonisation and development of roads, agriculture and industry, while boosting the economy and raising living standards. Besides direct government investment there are many factors that influence the spatio-temporal development of roads (Geist & Lambin 2002, Koorey 2009). However, despite the complex socio-economic drivers behind road development our results imply that the emergent process at larger spatial scales can be approximated by a simple logistic growth and dispersal process. These findings support those of Ahmed *et al.* (2013) who, using empirical observations, determined that road development displayed a logistic growth pattern through time, although here we also provide probabilistic estimates of the rate and asymptote of that process. Unlike Ahmed *et al.* (2013) we also find evidence for neighbourhood effects and characterise the nature of those effects in the form of simple functional forms.

We find little difference in model predictive performance when using different neighbourhood effects models so we avoid interpreting why we obtain subtle differences in predictive performance for those different dispersal functional forms. However all neighbourhood effects models support neighbourhood influences that are unidirectional – from the region of high road density to the region of low density. Together this implies that the spread of road development at large spatial scales can be characterised as a growth and diffusion type process. This is not surprising, however characterising regional road development in this way enables estimation of the rate and direction of road spread across the region. The development of roads away from the arc-of-deforestation, towards the centre of the Amazon reflects the economic activities of the area, where initial roads grant access to extractive industry and colonisers who expand the network with unofficial roads to increase access and transport products (Fearnside 2008, Perz *et al.* 2008). Over time, as more timber and land resources are exhausted, roads are built to access forest further from the arc-of-deforestation. Moving away from the arc-of-deforestation also helps integrate remote Amazon regions; an aim of initiatives such as Avanca Brazil (Fearnside 2008). The estimated speeds of road development are somewhat alarming, with both wave and neighbourhood effects models estimating complete road coverage across the Amazon within 60 years. However these projections should be interpreted as what would happen if the inferred pattern of road development between 2004 and 2007 were to continue for the next 60 years. Clearly multiple factors will influence the rate of road development to prevent this from happening such as, for example, barriers to road development as illustrated in Figure 6.8.

Our best fitted models explain around 20% of the variation in the data, clearly leaving room for improving predictions by incorporating more mechanisms. This would be a natural area for future work and could be conducted by extending the models we developed here within our parameter inference methodology (our source code can be downloaded from http://research.microsoft.com/en-us/downloads/b0cf61db-3c9d-4154-b5c1-5e5f72655185/default.aspx, Appendix A). One natural direction would be to infer the effects of different barrier types on the rate of road development. We performed a preliminary investigation into this by inferring parameters to a neighbourhood effects model in which we reduced the maximum road density in a given cell (parameter K) in proportion to the area of that region that would act as a barrier to road development (rivers and protected areas). However this led to no detectable improvement in the model fit or the inferred parameters and subsequent detailed investigation of the empirical data indicated many areas where roads developed in high density along river edges and extended into protected areas. This highlights an obvious area for future work; incorporating the relationships between road development and protected areas and barriers could enable more realistic future projections.

Predictive models of land use change are important tools for many other analyses including, climate change, carbon and biodiversity modelling. Additionally land use change models provide important information for decision makers. The importance of incorporating information on roads to improve the predictive accuracy of land use change models has been repeatedly demonstrated (Geist & Lambin 2002, Fearnside 2008). Consequently, roads have been found to be one of the most commonly used inputs in land use models in the Amazon, with a recent review reporting 24 out of 35 published studies utilise information on roads as inputs to models (Rosa *et al.* 2014). However, predicting road network development remains a challenge in predicting future deforestation because of uncertainty over how best to model the scale and interactions of the interdependent factors involved (Barlow *et al.* 2011). One natural extension of our work here would be to couple simple road development models, such as those developed here, with other simple models of dynamical processes influencing deforestation. To date, dynamic deforestation models and road development models have principally been developed and studied independently, with only two modelling platforms combining deforestation and dynamic roads (Jiang 2007, Soares-Filho *et al.* 2006).

Developing simple models of their coupled dynamics would enable deeper insights into their dynamical interdependencies and the critical factors influencing the rate and nature of their spatiotemporal dynamics. Our study provides a way forward for developing more accurate predictive road development models towards this end.

Chapter 7: **Process based modelling of road development in the Amazon: A proof of concept**

7.1. Abstract

Timber is a huge global market and a huge industry in Brazil, with the Amazon region producing an estimated 24 to 28 million m^3 of roundwood timber annually, generating a revenue of US$ 2.5 billion. Much of the road network in frontier regions are first developed by logging enterprises, with much of this development being unplanned. Road networks are growing at rapid rates in the Amazon (17,000 km per year on average), with much of this growth occurring along the arc-of-deforestation. Road location is a key input factor in many LULC (land use land cover) models focussed in the Amazon region, however the majority of these models treat road networks as static phenomena. Currently there are four spatially explicit models of road network expansion in the Amazon (Soares-Filho *et al.* 2006, Jiang 2007, Arima *et al.* 2008, Walker *et al.* 2013). We present a new process based model of road network development and while our model has not accurately replicated spatial patterns, we believe the model approach has sound rationale and represents several advantages over existing models. Firstly we estimate uncertainty, which other road models omit. Second, the rate of road expansion in our model was estimated from a large empirical data set. Third, other models which have been validated, have only been validated with spatial congruence, which is a ‘generous’ method that only considers ‘true positives’. We present more rigorous validation techniques which should be applied to road models. Fourth, our model uses a detailed map of timber value derived from economically important species, while other models have used total wood density as a proxy for value. Finally, we have modelled over larger spatial scales and larger networks than existing validated models.

7.2. Introduction

Roads can be a cause or a consequence of economic activity. As a cause; government initiatives are used to connect areas and reduce transport costs (effort and finance), to encourage economic activity and increase living standards. As a consequence; roads are built with the express purpose of accessing resources, for example logging roads that are built to access and transport timber. In many emerging economies, road building is vital for stimulating and maintaining economic growth (Andersen & Reis 1997). In Brazil infrastructure initiatives have been used since the 1970s to this end (Carvalho *et al.* 2002, Alves 2002, Kirby *et al.* 2006, Ahmed *et al.* 2013), indeed such initiatives are still in use today in the region, for example the Initiative for the Integration of the Regional Infrastructure of South America (IIRSA) project (Killeen 2005). The building of any road comes with consequences, generally positive for people (Calderon & Serven 2004, Straub 2008, Perz *et al.* 2012) and negative for the environment (Forman 1998, Spellerberg 2002, Coffin 2007). Roads that extend into forest frontiers are particularly damaging because they open access to areas that were previously ‘protected’ by their inaccessibility (Armenteras *et al.* 2006). Here we focus on logging roads built by loggers to access timber in the Brazilian Amazon.

Timber is a huge global market and a large industry in Brazil (Sierra 2001, Dauvergne and Lister 2012). The majority (90%) of timber in Brazil is sourced from natural forests (Sierra 2001, Matricardi *et al.* 2005). Approximately 350 Amazonian tree species are commercially harvested, producing an estimated 24 to 28 million m^3 of roundwood timber annually, generating a revenue of US$ 2.5 billion (Verissimo& Cochrane 2003, Merry & Amacher 2005, Arima *et al.* 2005). Logging operations in Brazil really began in the 1970s and timber was often close to mills with loggers usually travelling a few kilometres from mills to access timber. By the mid-1990s loggers regularly travelled over 100 km to access desirable timber (Johns *et al.* 1996). This has inevitably led to extensive road network development by loggers to access timber stands and transport timber to mills. In addition, many new mills have been set up deeper into the forest frontier along the new networks (Merry & Amacher 2005) and in many frontier regions of Brazil it is logging activity that is the main cause of road construction. One study found two thirds of roads surveyed near Tailandia, Pará, were built by loggers, often in exchange for logging rights on the land (Uhl *et al.* 1991). Interestingly, fewer roads are built under planned (or reduced impact) logging operations compared to unplanned (and illegal) operations extracting the same volume of timber, for example Periera *et al.* (2001) found that roads covered 1.2% of the harvest area in an unplanned operation compared to just 0.6% in the planned harvest area, two years later the planned area had 1% road coverage and the unplanned area had 2%. The vast majority of logging, approximately 95%, is unplanned which causes more damage than planned logging (Johns *et al.* 1996, Pereira *et al.* 2001, Verissimo *et al.* 2002).

Even when forests are selectively logged with little environmental damage,many areas that are logged are often deforested within a few years (Asner *et al.* 2006), primarily because access is granted to agriculturalists, land prospectors, and colonists who utilise the roads built by loggers and cause deforestation and degradation (Fearnside 2007, Laurance *et al.* 2004). Logged regions are also at greater risk of further forest loss through fire risk caused by edges created by roads (Broadbent *et al.* 2008, Nepstad *et al.* 1999, Nepstad *et al.* 2001, Uriarte *et al.* 2012). This makes the road network a key factor in deforestation patterns, with studies showing that roads and deforestation are closely linked (Chomitz & Gray 1996, Laurance *et al.* 2001, Perz *et al.* 2007, Laurance *et al.* 2009, Caldas *et al.* 2010, Southworth *et al.* 2011). Consequently, roads have been found to be one of the most commonly used inputs in land use land cover (LULC) change models in the Amazon, with roads determining the accessibility of land and the cost of transportation which in turn determines the viability of land use change (e.g. Messina & Walsh 2001, Soares-Filho *et al.* 2004, Lapola *et al.* 2010, Maeda *et al.* 2011). The importance of roads as an input for LULC change modelling has been repeatedly demonstrated, with roads being the single strongest predictor of spatial patterns of deforestation (Geist & Lambin 2002, Pfaff *et al*, 2007, Fearnside 2008).

Modelling the expansion of road networks is 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). The many difficulties associated with predicting this largely anthropogenic phenomenon that is subject to many idiosyncratic events possibly explains why LULC models treat it as a static phenomenon. Roads in the Amazon region are a dynamic, spatially explicit phenomenon, which have been growing at rapid rates; between 2004 and 2007 17,000 km of new roads were added per year (Brandão and Souza 2006, Ahmed *et al.* 2013). The majority of this expansion was concentrated around the arc-of-deforestation (Ahmed *et al.* 2013), a key area of land use change, and yet none of this expansion was taken into account by most Amazonian LULC models. This maltreatment of a key deforestation predictor could have serious repercussions on the efficacy of LULC models. While there are several modelling frameworks available to predict the development of road networks that have been used in LULC models (Messina & Walsh 2001, Soares-Filho *et al.* 2004, Soares-Filho *et al.* 2006, Lapola *et al.* 2010), there is no peer-reviewed literature presenting these road models, nor any numerical validations of the road model predictions. While it is 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. Thus it is important for LULC models that rigorous road models are developed.

The location of a road depends on two main considerations: (1) where the road should go, i.e. where does it start from and where is its destination; and (2) constraints on the alignment of the road that impact its feasibility and/or cost, such as rivers, mountains and human land uses. Every existing road model based in the Amazon uses these two considerations in some way. Currently there are four spatially explicit models of road network expansion in the Amazon (Soares-Filho *et al.* 2006,Jiang 2007,Arimaet al. 2008, Walker *et al.* 2013), all of which use least-cost paths to determine the route a new road might take. Only two of the four models have been formally tested and validated against empirical data (Arima *et al.*2008, Walker *et al.* 2013), but these models have not yet been incorporated into regional LULC models.

The first of the four road models of road expansion 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 are used to determine the destination 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). This model provides deterministic roads and single map outputs, which have not been validated against real road development patterns. Primary roads are not predicted, with the model instead relying on known government planned roads to form the initial network from which to develop roads in frontier regions. However, when roads develop into frontier regions it is often logging roads, rather than government roads, that cross the threshold first.

The second road model, IDRISI's road extension module (Jiang, 2007), is based on similar principles as DINAMICA but produces a hierarchal road network by allowing different spatial structures for primary, secondary and tertiary roads. It incorporates the cost of converting different LULC types into a road into 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. While this model allows different spatial structures for different orders of roads to be incorporated, it is deterministic, producing a 'rigid' network based on a combination of numerical rules modulated by the least cost path algorithm.

The third road building model was developed in two stages and attempted to recreate the road building decisions made by the logging industry (Arima *et al.*2005, Arima *et al.* 2008). In the original model, Arima *et al.* (2005) predicted both destination determinate roads (where road destinations are selected and a road is built from the chosen destination to the existing road network) and destination indeterminate roads (where roads simply grow from the existing network with no fixed destinations). In the Arima *et al.* (2008) model only destination determinate roads were considered. This later model recreated the road building decisions made by the logging industry (Arima *et al.* 2008), 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. Approximately 150 km of roads were modelled, with the modelling carried out in a small (46 x 38 km) area in ‘Terra do meio’, Para, originally a major logging frontier until it was placed under protection in 2004 (Walker *et al.* 2013).

The final road model (Walker *et al.* 2013) also focussed in the ‘Terra do meio’ region, and the networks used in the study were specifically selected to reflect autonomous dendritic networks that could be attributed to single logging operations. Two small study sites of approximately 45x35 km and 20x40 km were used and approximately 140 km of roads were modelled. Here logging road development was formalised into a graph theory frame work where logging sites are ‘nodes’ and connected with roads known as ‘edges’. Many different ‘graphs’ (i.e. networks of edges connecting nodes) are compared and the graph that is ‘optimal’ is considered the road network, in this case for example, the graph that had the highest profit. Profit was calculated by taking the cost of all roads (a combination of distance between nodes and topography) from the total revenue from the logging sites (where revenues are proportional to total wood density). All destination nodes in the model were predefined. This means that for any given optimisation problem there is only one or a small number of equivalent solutions, i.e. the model is deterministic. In terms of time scales, the model assumes a constant development of roads of between 5 and 10 km per week over a 5 year period (1996-2001), based on surveys conducted in 2003/4 with various stakeholders. Validation of the model showed between 56% and 75% of predicted roads fell within a 1026m buffer of observed roads.

These four models produce deterministic, single map outputs. Those that have been validated (Arima *et al.* 2008, Walker *et al.* 2013) are based in a relatively small study site (not even 100 km x 100 km), modelling a relatively small road network (no more than 200 km). Where details of temporal calibration are evident, this has been based on 'expert knowledge' (Walker *et al.* 2013). The rate of development in the models, DINAMICA and IDRISI, can be calibrated based on past road data that the user possesses, and they are designed to be implemented at any spatial scale thus are compatible with the LULC models. However they have not been validated against real road data. The models of Arima *et al.* (2008) and Walker *et al.* (2013) though validated, have not been implemented on the same spatio-temporal scales as LULC models, making them potentially unsuited to LULC applications.

We present a process based model following concepts presented by Soares-Filho *et al.* (2006), Arima *et al.* (2008) and Walker *et al.* (2013); using a combination of land attractiveness to determine road destinations, and a least cost path algorithm to determine the alignment of new roads. Our modelling framework has several advantages over existing models, chief among them being that it quantifies the uncertainty around predictions, which no existing models do. We use a more detailed measure of land attractiveness to determine the destination of roads, and a friction map that takes more constraining variables into accountto quantify alignment constraints. We calibrate the model using empirical data derived from a space-for-time substitution from 443 municipalities in the Brazilian Amazon (Ahmed *et al.* 2013). Further, we use annual time steps, a larger study site (185 x 185 km), and attempt to model over 700 km of roads. We focus on logging roads in the Brazilian Amazon because many roads are developed unofficially by loggers, logging is the primary driver of road development in the region (Arima *et al.* 2005), timber is a valuable and widespread commodity in the region, and the Amazon is a highly vulnerable and diverse ecosystem under threat from logging operations.

7.3. Methods

7.3.1. Study sites and roads

A Landsat location (path/row ID: 231/065) was chosen as an initial site in which to model road networks. We chose a low density location, as we know high density locations experience little change in road network through time (Ahmed *et al.* 2013). Road densities from Landsat locations covering the Brazilian Amazon were calculated and divided in ArcGIS into three density classes based on natural 'Jenks' breaks, low (0.00-0.05 km/km^2), medium (0.05-0.13 km/km^2) and high (0.13-0.23 km/km^2) density scenes. From within these classes the scene was selected to represent a region of low (0.02 km/km^2) road density. The road network was manually digitised based on annual Landsat 5 TM images from 2000 to 2008, following the methods of Brandão & Souza (2006). There are many automated approaches to digitising road networks (Mena 2003, Brandão & Souza 2006, Li & Briggs 2009, Movaghati *et al.* 2010), but these are typically less accurate than manual digitisation (Li & Briggs 2009). Validation of the images by Ahmed *et al.* (2013) found an average spatial congruence of 82.5% within a 200m buffer of road maps independently generated in the same Landsat scene by IMAZON, a Brazilian research institute that maps unofficial roads among other projects (IMAZON 2011). The road map for the year 2000 served as an initial road network upon which to base the model and the subsequent road maps were used for validation.

7.3.2. Process-based model of road development

The model was written in python 2.5.2, with the numpy 1.5.1 and dbfpy libraries (available from: http://sourceforge.net/projects/numpy/files/NumPy/ and http://sourceforge.net/projects/dbfpy/files/ respectively), and utilises ArcGIS 9.3 tool boxes. The model code is presented in Appendix C and a flow diagram of processes is in Figure 7.1. All spatial inputs were on 250m grids projected in SAD 1969 continental projection. The road network development model is based on the simple assumption that loggers building roads want to make a profit by maximising revenue and minimising costs (Arima *et al.* 2005, Arima *et al.* 2008, Merry *et al.* 2009, Walker *et al.* 2013). We simplify this process to consider revenue as being determined by the value of timber harvested, and costs as being determined by the amount of new road to be constructed. In summary, the steps taken in the model are: (1) Destinations for new roads are determined according to the distribution of potential timber revenue, weighted by the costs of building roads to those destinations; (2) the amount of new road to be built in a time step is estimated from observations of contemporary road network growth rates across the Brazilian Amazon; (3) a new road is constructed to the selected road destination along a least cost path; (4) new destinations are selected and roads constructed until the expected amount of new road has been constructed in that time step; (5) a spreading dye model simulates logging and deforestation around the new road based on costs of accessing timber via skidder tracks; (6) steps 1-5 are repeated for each time step in the simulation.

In this model three factors need to be determined: (1) road destination;(2) road alignment, or the route the new road takes; and (3) the amount of road to be built in each annual time step. Road destinations were determined from a map showing the spatial distribution of potential revenue obtainable per hectare from timber extraction across the Amazon (Ahmed *et al.* 2012). Wood density has been used before to determine road destinations in road construction models (see Arima *et al.* 2008, Walker *et al.* 2013), however not all tree species are equally valuable therefore overall density is not the best measure of revenue for an economic model. We combined RADAMBRAZIL tree survey point data with timber value (US$) from ITTO, for 11 commercially valuable timber genera and used krigging to interpolate values across the region (Ahmed *et al.* 2012). This layer determines the destination of new roads, with destinations being within a buffer that measures the total predicted road length to be added for that time step. The model then selects out the top 5% of valuable cells to pass onto a weighted sampling process, and then picks a location to build the road to, based on a local estimate of revenue divided by cost using a circular weighted average. We derive the cost from a friction map, which determines how expensive (in relative terms) it is to build through a given location (see below for details). This determines the 'profit' available from a given stand if a road were to be built to it. Groups of adjacent grid cells that have similar profit are made into potential destination timber stands. The model picks a random timber stand to build the road to with the choice weighted by the skidder extraction ratio. The extraction ratio is taken to be the cost of extraction by skidder; 0.25 of the cost divided by revenue ratio for cells around the chosen location (this ratio is deliberately inverted from the first ratio). The relative cost of building skidder trails compared to building a road of the same length we have assumed to be ¼ the cost, thus the cost layer is multiplied by 0.25 for this ratio. The accumulated distance in cost/revenue units from the road destination (the centroid of the chosen location) is calculated across the timber stand. The lowest values across this accumulated surface show cells that either have close proximity to the road end (few cells to accumulate to get the cost there) or are higher value/low cost cells (lower value cells to accumulate) or both. The model then selects a number of cells capturing the lowest quantile of this surface that corresponds to the target concession size based on an ink diffusion model.

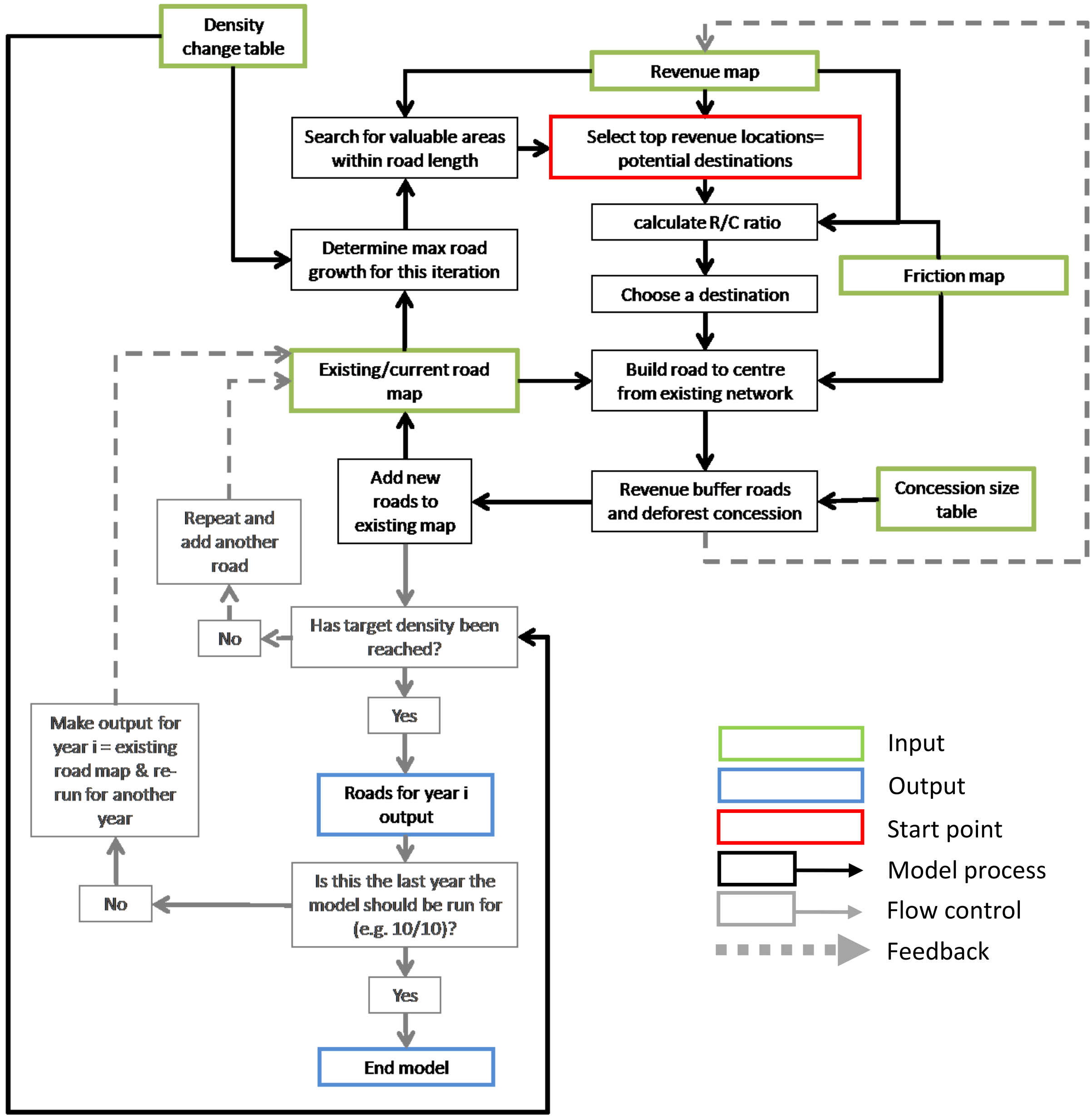


Figure 7.1. Model process schematic (for model code see Appendix C).

Road alignment was predicted using ArcGIS's least cost paths tool and a friction map representing the relative costs of road building per pixel. Here, we assume that the cost of materials and labour per kilometre of road is constant and can be ignored, allowing us to focus on the cost in terms of difficulty of building a road through constraints as represented inthe friction map. Factors that can constrain or facilitate the laying of the road fall broadly into four categories, topography, existing developments, hydrological features and ground conditions (Koorey 2009). The friction costs associated with each factor are described in Table 7.1:

(1) Topography is widely accepted as being a key determinant of road alignment, with it being more difficult and costly to build up a steep slope than along flat land (Liu & Sessions 1993, Soares-Filho 2006, Arima *et al.* 2008, Walker *et al.* 2013). We used a DEM (digital elevation model) of the Amazon Legal region as the basis of our friction layer. The map was derived from SRTM altitude data based on the change in altitude between two adjacent pixels. This layer formed the basis of the cost map, into which the other factors were weighted and multiplied in.

(2) Roads have been observed to develop three times slower in protected areas within the Amazon region (Barreto *et al*. 2006). Thus we multiplied the topography map with a map of protected areas using ArcGIS Map Algebra tool, where each protected area grid cell was assigned a value of three.

(3) Rivers are also known to influence road development, especially logging roads where investments in road paving and bridge building are minimised. We multiplied the rivers through our DEM with an assumed cost value of 5.

(4) The last factor we considered was ground type because it is hardest to build roads on very wet or waterlogged clay soils and on very loose sandy soils. A map of soil types for the Amazon from IMAZON identified nine soil types, ranging from waterlogged to very sandy. Friction values between 1 and 2 were assigned to the nine soil types, such that intermediate soil types were easiest to build on and extreme types more difficult.

Table 7.1. Factors and weights used to generate the friction map.

Friction layer inputs	**Multiplication factor**
Topography	1 (Base layer)
Rivers	5
Protected areas	3
Soil type	1-2
Waterlogged	2.00
Very clay	1.75
Mid-very clay	1.50
Mid-clay	1.25
Clay	1.00
Mid-soil	1.00
Sand-clay	1.25
Mid-sand	1.50
Sand	1.75

The higher the value on the friction map the harder it is for a road to pass through that area. Road alignment from the existing road network to the new road destination was determined using the Dijkstra least cost algorithm in ArcGIS.

The amount of new road to build in each time step was determined from an analysis of road network growth rates across the Brazilian Amazon (Ahmed *et al.* 2013). We constructed a table of observed initial road densities in 2004 and annualised changes in road density over the period 2004-2007 for each of the 443 municipalities of the Brazilian Amazon (Figure 5.2). Here, we selected the subset of municipalities that had an initial density that fell within ± 0.05 $km.km^{-2}$of the road density in our modelled landscape, and selected at random the annualised change in road density from one of those municipalities to be the target change in road density in the model for that time step (see simulations of road network expansion described in Ahmed *et al.* 2013 for more details). Because it is very unlikely that any one (or set of) new road built from the existing road network to the chosen destinations will exactly match the length of new road to be added in a given time step, we introduced an arbitrarily chosen tolerance of ±20% to the amount of new road, ensuring that only 'complete' roads are built each time step. New roads were added to the existing road map sequentially until the targeted increase in road network density had been reached, after which the model moved onto the next time step.

Once a road has been constructed to a selected timber stand, an area is 'logged out' of the revenue map and is considered to have zero potential revenue the following time step. Logging areas were intended to represent approximately the size of an annual timber concession in the region. We obtained a list of observed concession sizes taken from the Brazilian government forestry service (SFB 2013) and divided those areas by 40, as the average concession lasts 40 years, to annualise the values. From this table we selected one concession at random to determine the concession size to be logged at each road destination. The spatial pattern of logging was modelled using an ink diffusion model, beginning from the road destination selected above, spreading out to an area that corresponds to the target concession size, considering that area to have been logged. Furthermore, we applied a 2 km buffer along the length of the newly developed road that we assumed would be deforested. This represents a conservative estimate of the distance to which roads influence deforestation (Brandão *et al.* 2007, Southworth *et al.* 2011). Logged and deforested areas were assigned a value of zero potential revenue for all future time steps in the model iteration, ensuring no additional roads would be built into those areas.

The model we developed incorporates uncertainty in several manners. First, we used empirical observations to determine the rate of road growth in any given time step. This approach allows ‘jump-start’ development (Ahmed *et al.* 2013) and it inherently takes distributions of potential changes into account. The random selection of an annual logging concession size for each destination varies the amount and location of suitable destinations in each time step for each iteration of the model. Further stochasticity is introduced in the selection of road destinations based on random weighted selection of potential logging sites, and the area logged by the ink diffusion model. These sources of stochasticity make it possible to quantify the uncertainty around our model predictions. We simulated road network expansion in a Landsat scene (path/row ID: 231/065), using the road map in the year 2000 as the initial road network and predictions were made using annual time steps for the period 2001-2008, coinciding with the years for which we had data to validate our predictions against. Our model was stochastic, meaning that each iteration generated a different result, so we iterated the model 100 times. We quantified the likelihood of a predicted road occurring within any given cell as being the proportion of model iterations in which a road was predicted to occur in each cell.

7.3.3. Model validation

We compared the average predicted road network length with the observed length for each year (2000-2008), testing the models ability to predict the amount of road growth and the rate of growth. We tested whether the model predictions for 2008 performed better or worse than random using a ROC curves and AUC (area under the curve) measures. This was calculated using the probability of a predicted road being placed in a given cell and whether a road actually occurred in a given cell by 2008, in the ROCR R package (Sing *et al.* 2005), for each of the 100 iterations. This represents an extremely stringent test of model accuracy, as 'near-misses' are treated exactly the same as 'far-misses' (Pontius *et al.* 2002, Pontius *et al.* 2004). Thus we also assessed spatial congruence between the observed road network in 2008 and each of the predicted 2008 road maps. We determined the proportion of predicted roads that fell within a series of set distances from observed roads (250, 500, 750, 1000, 1500, 2000, …,, 10000 m). We conducted this test twice, once using buffers around all roads present in 2008 and once using buffers around just the new roads created post-2000. For each of the 100 model iterations, we generated a neutral prediction against which to compare our model predictions by calculating the proportion of cells occupied by a road in 2008 and randomly assigning the presence of a road to the same proportion of cells. We then applied the same set of buffer distances used in the spatial congruence analysis around the observed roads and calculated the length of random roads falling within each buffer. Two-sample t-tests were used to determine if the proportion of predicted roads falling within each buffer was significantly higher than the proportion of neutral roads. Further, we used binomial GLMs to determine any statistical relationship between the predicted probability of a road occurring in a grid cell vs. whether or not a road was actually observed. Each GLM was conducted on a 10% subset of the data to reduce the degrees of freedom.

We also validated our model using an emergent metric of a road network termed roadless volume (Watts *et al.* 2007), which we calculated for the observed and all 100 predicted road networks. Distance to nearest road was calculated on a 250 m grid using ArcGIS Euclidian distance tool, summed together and divided by the total area of the model region. Comparing the roadless volume of predicted networks and the observed network gives an indication of how well we predicted the overall network spread and road arrangement in terms of area 'undisturbed' by roads.

Lastly, to determine if the destinations of roads were being correctly predicted we extracted the end point of all observed roads (2008) and all predicted roads in the 100 model iterations. Kernel density was calculated in ArcGIS using the predicted end points, and observed end points were overlaid. To assess how well predicted destinations matched observed destinations, we calculated the mean distance between every observed point with every destination point and the mean nearest distance between every observed point and its closest predicted destination. We then repeated this using randomly placed destination points to provide a neutral comparison. Preliminary observations suggested that model performance was strongly influenced by the ability to predict destination points, so to determine if the model would correctly align roads to known destinations we ran the model with the observed 2008 destinations and carried out spatial congruence analysis on the resulting paths with those roads observed from 2008.

7.4. Results

The study site included a total of 594,185 cells, of which 72,535 (12 %) were predicted to have a road in at least one of the model iterations. Of those cells predicted to contain a road, only 88 had a probability of greater than 0.7, the majority (63,292) had a probability of less than 0.1, highlighting the low certainty with which our model was predicting the location of new roads (Figure 7.2). However, the binomial GLM's indicated that in general cells with larger predicted probabilities have an increased probability of new roads occurring (mean slope=3.2 (CI=3.1-3.6), mean SE=1.00 (CI= 0.98-1.03), df=59418, mean p=0.031 (CI=0.031-0.034)). A total of 616 km of roads were added to the observed network between 2000 and 2008, whereas our model predicted an average increase of 710 km (95% CI= 658-761km). Thus the model over-predicted the amount of road added to the network each year (Figure 7.3). Over prediction ranged upto a maximum of 777 km by 2008 in one model iteration, with an average over prediction of 93 km (95% CI= 42-144 km) by 2008 across all 100 iterations. Over the eight year period, across all iterations, we found an average mean annual over prediction of 21% (range 11% (2001) -33% (2004)). While we over-predicted the absolute amount of roads in the network, the rate of road network expansion was almost perfectly predicted; regressions of observed and average predicted road growth as a function of year had slopes that were not significantly different (p=0.55) (Figure 7.3) (observed rate: 83 km/year, SE= 6.09, df=6, p<0.001, r^2=0.96; predicted rate: 86 km/year, SE=2.15, df=6, p<0.001, r^2=0.99).Predictions decreased in certainty through time, with a difference in upper and lower 95% confidence intervals around the mean prediction of 38km in 2001, increasing to 103km in 2008.

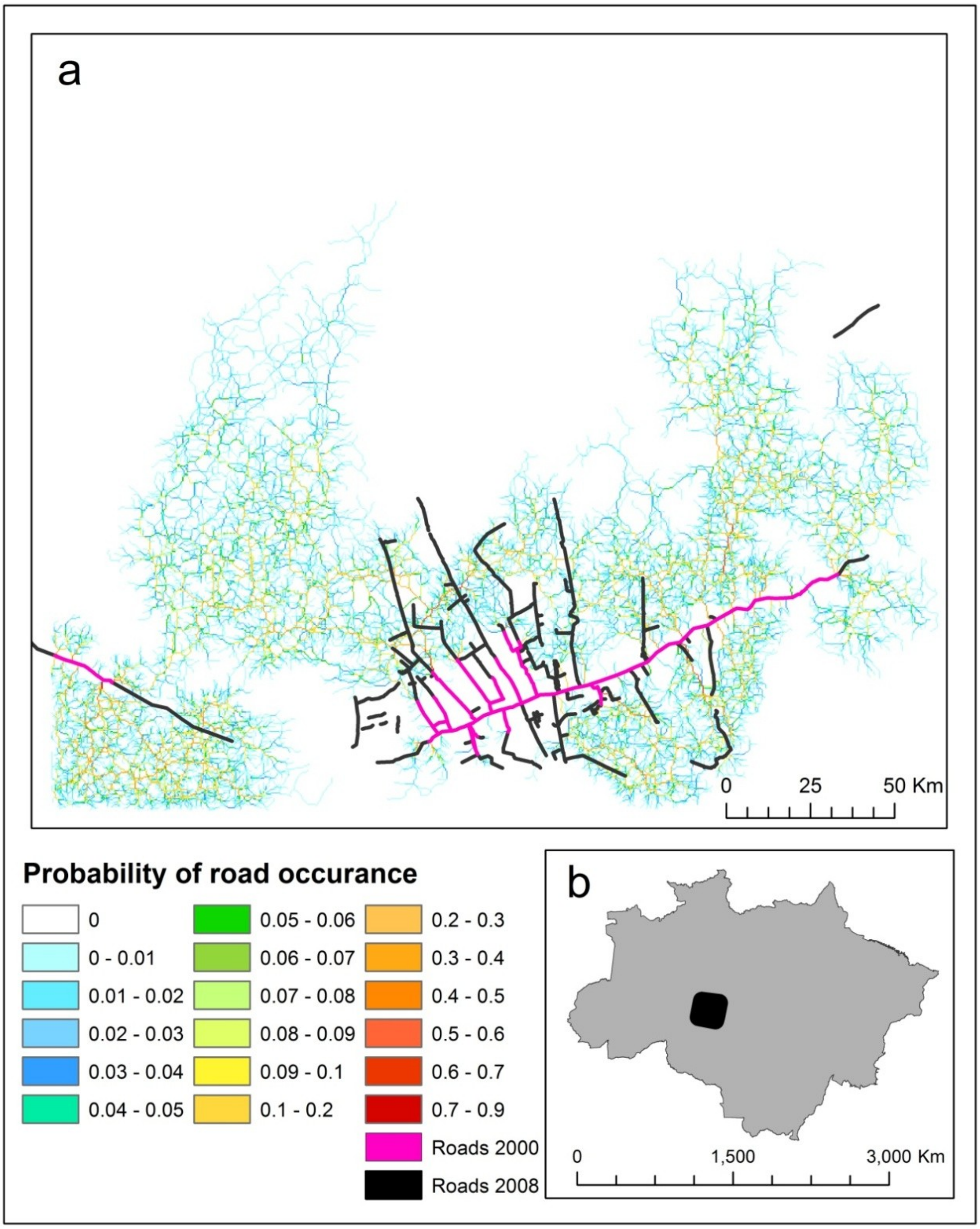


Figure 7.2. Map of probability of a road occurring in any given cell. The observed road networks for 2000 and 2008 are shown for reference.

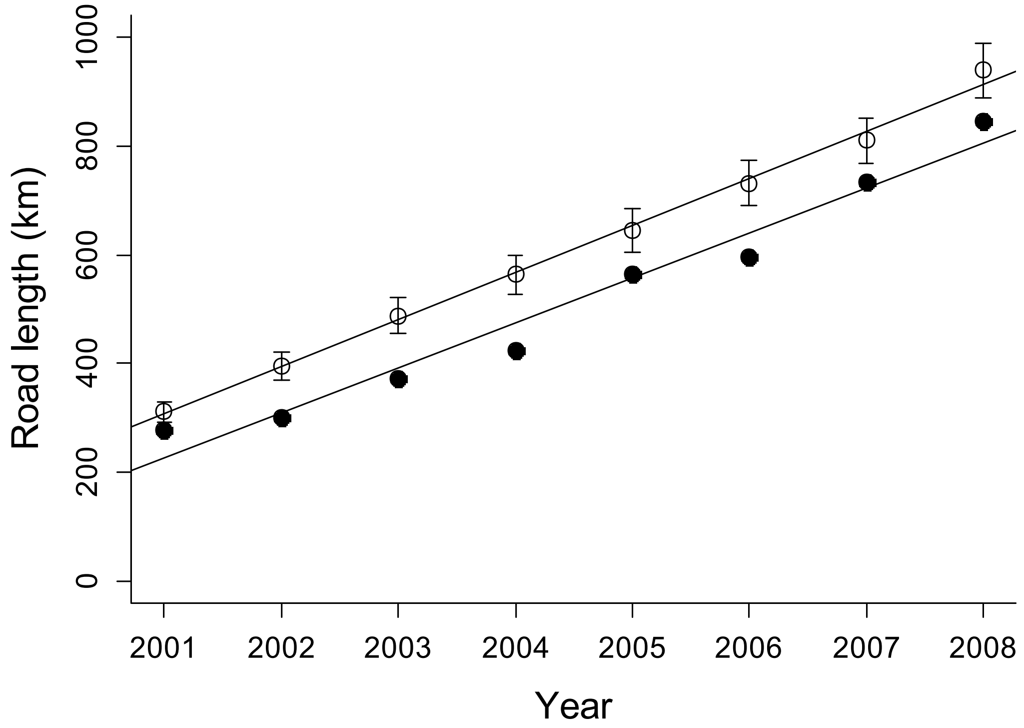


Figure 7.3. Annual observed and predicted road lengths. Mean predicted road lengths are open circles with 95% confidence interval error bars, observed lengths are closed circles. Fitted lines are derived from regression models

ROC curves indicated the model predicted the pattern of road occurrence did not perform better than random with an average AUC value of 0.5033 (CI=0.5032-0.5035), this impression is reinforced by qualitative visual interpretation of the predicted and observed networks (Figure 7.2). Spatial congruence, however, found an average overlap between predicted and observed roads of 80% at a 5 km buffer, if all roads are considered (i.e. the total road network), for comparison Arima *et al.* (2008) found a 90% overlap at 5 km. If we remove the road network from the year 2000 and only consider the change in the road network, ~70% of predicted roads still fall within 5 km of observed new roads. Spatial congruence predictably declined with decreasing buffer sizes (Figure 7.4), but individual t-tests at all buffer sizes showed that the spatial congruence of predicted roads was significantly higher than the spatial congruence of random roads for both the total road network and for change in road network ($P<0.05$ in all cases).

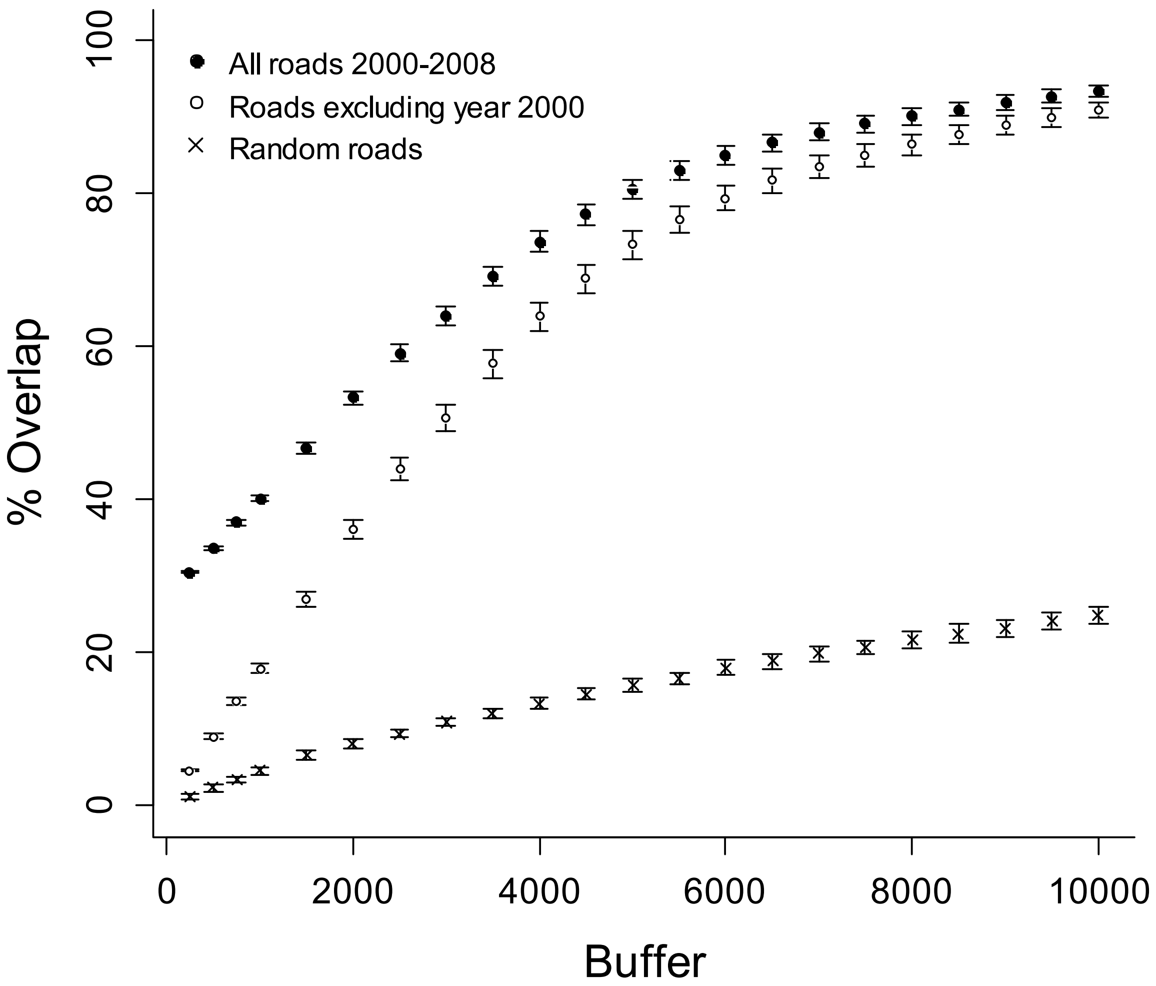


Figure 7.4. Spatial congruence showing mean percent overlap between predicted and observed roads (±95% CI error bars). Closed points show all roads from 2000-2008, open points show roads from 2001-2008 (i.e. change in road network), and crosses show the mean percent overlap between random roads and the observed network.

Roadless volume in 2008 for the study site was 164, which was slightly higher than the average predicted roadless volume of 146 (95% CI= 139–152, range = 85– 235). While the range of roadless volume predictions overlaps with the observed roadless volume, we fail to capture it within our 95% confidence interval. Further a one sample t-test shows that the observed and predicted roadless volumes are significantly different ($t=-5.6$, $df=99$, $mu=164$ $p<0.05$).

A total of 308 road destination points were extracted from the observed data and a total of 45,752 destination points were recorded across all 100 model iterations, averaging 456 (95% CI=430-483) destinations per iteration. Our model, then, predicted the occurrence of ~33% more road destinations than were observed, which is in line with the over-estimates of road length described above. On average, predicted destinations fell 57 km (95% CI=54-60 km) from observed road destinations, whereas randomly selected destination points fell 91km (CI=88-94 km) from observed destinations, which is significantly further away ($t=48.3$, $df=614$, $p<0.05$). Kernal density of predicted points revealed three key areas of predicted destinations where there were no observed destinations (Figure 7.5). To determine if the model would build the observed roads if it was given the observed destinations we ran the model with the observed 2008 destinations (Figure 7.6). We found that the model performed much better with known destinations, finding a spatial congruence of 98% of predicted roads falling within 2 km of the observed network.

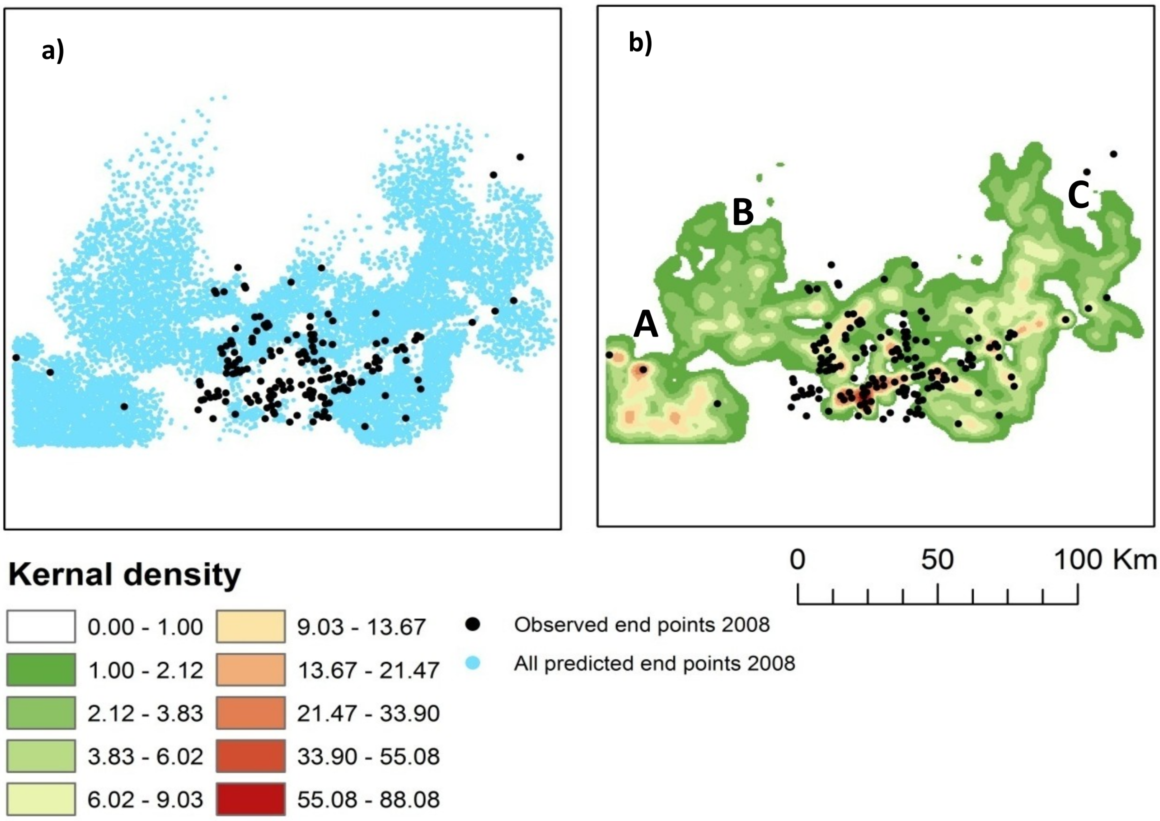


Figure 7.5. Observed vs. predicted road destinations. a) Observed road end points (black circles) surrounded by all predicted end points across 100 model runs. b) predicted end point kernel density showing 3 main areas (A, B, C) that do not match the observed road destinations.

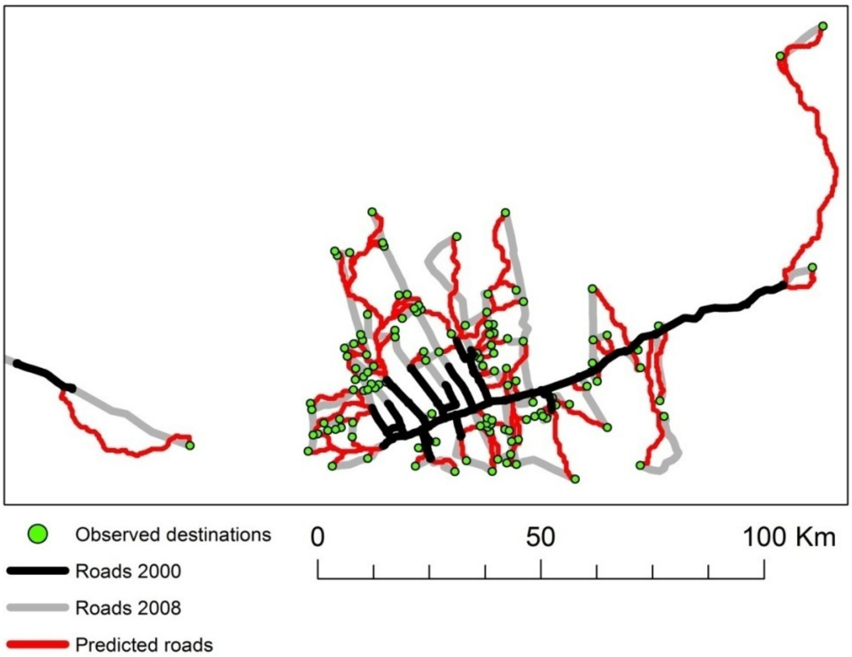


Figure 7.6. Model predicted road paths from 2000 to 2008, given observed 2008 road destinations.

7.5. Discussion

We have developed and presented a process based model of road network expansion in the Brazilian Amazon, explicitly linking the construction of new roads to the expansion of logging and validating model predictions over an eight year period. While there is much improvement to be made in terms of prediction accuracy, the model serves as a reasonable starting point for future development.

Our model over predicted both the total amount of road added to the network and the number of new roads to be constructed. This is most likely associated with 'jump start' development of road networks, where low road density areas such as our study area can experience rapid growth in a short period of time (Ahmed *et al.* 2013). This process did not occur in our study area, yet our model allowed for that possibility and consequently over-predicted the total amount of new roads that were constructed. The rate of road network expansion, however, was almost perfectly predicted, suggesting that the method for determining rates of new road to build is a reliable one. We suggest that this method for calculating the amount of new road could potentially work better in mid and high density areas that would not be affected by jump start scenarios, and that in these locations any over-prediction of road network growth rates would be reduced.

While we had high levels of spatial congruence at large buffer sizes and were able to demonstrate that in terms of spatial congruence the model performs significantly better than a random neutral model, ROC analyses showed that the model does not really perform better than random. This apparent contradiction probably arises because spatial congruence only considers true positive results, when ROC also takes into account false positive results. The reliance on true positives ensures that when the amount of roads are over-predicted, there is a greater chance of overlap between observed and predicted roads, and this will falsely inflate metrics of model performance. Furthermore, spatial congruence may perform better than random because of the clustered nature of this particular road network. In our model area, roads primarily occurred in the bottom half of the area and predicted roads spread out from this starting point, whereas our neutral comparison randomly distributed roads over the entire study area. If the initial road network were more centrally positioned within the study area, it is possible that the very significant difference between spatial congruence with predicted and random roads would be reduced.

The stochastic nature of our model resulted in a wide spatial spread of predicted road destinations (7.5), which led to greater uncertainty in the road predictions. Some of this variation may have been reduced by iterating the model more than 100 times. However, the fact that the predicted destinations fell much further away from the initial road map that the observed destinations suggests that more iterations would not have greatly increased the accuracy of predictions. The size of the deforestation buffer applied around new roads could be the reason that road destinations were 'pushed' progressively further away from the observed network, a smaller buffer in future could help reduce this error. In turn, the wide spatial spread of predicted road destinations was responsible for the roadless volume predictions being lower than observed. The majority of our predicted destinations fell far outside the distribution of observed road end points, thus causing a greater proportion of the study area to be 'impacted' by roads and resulting in a lower roadless volume. When the model is run with known destinations the performance is considerably increased with 98%

spatial congruence within 2 km (8 grid cells). However, we still find discrepancies between observed and predicted roads. This is most likely because 1) we do not know the relative costs of the friction map inputs which affects alignment on the ground, 2) it is likely that some features that can influence building are not evident on 250 m GIS data, and 3) when there is little difference in cost between two road options a straight path is probably easier to build. Even with the discrepancies in path (Figure 7.6) we find a very high spatial congruence. Thus we can conclude that the model's overall poor performance is most likely related to destination selection rather than path building. We may be able to further improve the models path prediction if we use a calibrated friction map rather than one with weights based on conjecture.

Clearly the parameters used in our model require adjustment to ensure the model is better able to replicate observed patterns of road network expansion. The friction layer represents a key determinate of new road alignment, yet the factors included and weights allocated to those factors were largely based on assumptions and arbitrary decisions rather than on known costs. Future work will need to either work directly with the road-building contractors to obtain reliable friction estimates, or combine maps of observed road network changes with statistical techniques used to estimate the friction layers that determine dispersal patterns in biology (e.g. Richard & Armstrong 2010). Similarly, other model parameters such as the width of the deforestation buffer zone and the relative costs of building skidder trails vs. roads, were largely based on conjecture rather than hard facts, and highlight the paucity of information available on the true parameters associated with the logging and road construction industries. Another area of future improvement concerns validation. Here we validated the amount, rate and some measures of spatial match between observed and predicted networks. Spatial measures are particularly difficult to validate, given that spatial congruence has an over-reliance on true positives whereas ROC analyses do not allow for near-misses. Moreover, aside from the metric of roadless volume, we did not validate the emergent features of the network itself. Many network metrics have been developed and could be used to quantify the features of the road network as a whole (e.g. Freeman 1977, Theiler 1990, Sevtsuk 2010). These types of measures may be important to include because road building, especially as unofficial road building, is often stochastic with regard to exact location on the scale of individual roads. Measuring the network features with network metrics would allow us to describe the type of the network and determine if we are capturing the fundamental nature of development and resulting network, even if there are many ‘near-miss’ errors in the predictions.

It is clear that our model has not accurately replicated the patterns of road network expansion in the Brazilian Amazon, yet despite the poor overall predictions we believe the model approach has sound rationale and represents several advantages over the current suite of available models. Firstly we estimate uncertainty, which other road models omit, but which is clearly an important aspect of any predictions. We incorporate uncertainty from three main sources, the selection of road destinations, the amount of road built each year, and the amount of area logged around each destination (this affects the location of the next destination). The rate of road expansion was estimated from a large empirical data set, which leads to a logistic pattern of development. Other models appear to use a constant rate of development, which we know is unrealistic except over very short time scales (Ahmed *et al.* 2013). Our model’s spatial congruence is not dissimilar to results obtained by Arima *et al.* (2008) at the 90% congruence level, with 90% of predicted roads falling within 7km and

5km of observed roads respectively. Other validated models appear to have better predictions in general, however it is unclear whether these models function over multiple or single time steps. Further, we model over larger spatial scales and larger networks (~700 km vs. ~150 km of roads) than Arima et al (2008) and Waker *et al.* (2013). An advantage of our model over the most recently published road model (Walker *et al.* 2013) is that we allow road destinations to develop as the model runs, while Walker *et al.* (2013) must determine the location of destinations before the model begins. The static approach of Walker *et al.* limits the uncertainty of the model and means that changes to destinations are not allowed as the network develops even if a new area becomes an attractive potential destination. Our model uses improved inputs, a map of timber value derived from economically important species, while other models have used total wood density as a proxy for value. We also use a more complex friction map derived from, topography, protected areas, ground type and rivers while other models appear to only use topography, or a combination of topography and rivers, although they do acknowledge soils susceptible to flooding are important but not relevant to their area.

A key benefit to this modelling approach is that any factor that determines the attractiveness of land for a road destination can be incorporated into the revenue layer, not just timber value. For example, agriculture, mining, and oil and gas extraction all provide a strong impetus for the development of roads in the Amazon (Laurance *et al.* 2004, Fearnside 2007, Finer *et al.* 2008), and it would be possible to include maps of known mineral deposits and concession areas to the revenue layer to better estimate the destinations of new roads at large scales. Similarly, soil fertility and agricultural suitability measures could be included for roads built to exploit new agricultural areas. Planned development sites could also be incorporated, such as the locations of new dams and, obviously, any locations of known planned roads that will be built in the Amazon through official development plans such as IIRSA (Kileen 2005) and Pelt-Para (PELT-Pará 2012).

The model outputs from our framework can be parsed to LULC models that rely on road location as an input, alternatively, the model code could be incorporated as a module within a larger model, allowing direct incorporation of road predictions. While our model relies on logging and deforestation events to determine where not to build roads to, LULC models rely on the location of roads to determine where to deforest. By combining our model with a LULC model this feedback loop could be explored, potentially leading to new insights into both the dynamics of LULC and logging road network development.

Modelling how unofficial roads could develop from planned changes to the network would allow future decisions and scenarios to be better evaluated. In order to generate accurate road predictions much more work needs to be done at small scales (as presented here) and appropriate scaling up to larger regional areas needs to be carried out. The work presented here provides some advancement to this end.

Chapter 8: **Discussion**

8.1. Discussion

The main objectives of this thesis were (1) to investigate the ecological effects of roads, (2) to generate amazon-wide road models that could potentially be used to estimate the future of road development and their impacts in this globally important biome, and (3) to determine if these models would stand up to critical validations. The majority of the work presented here contributes to the modelling aspect of these aims, with some consideration of ecological effects. In Chapter 3 one of the many road induced ecological impacts presented in Chapter 2, namely avian biodiversity change related to road development, was considered. Two approaches to modelling road development have been presented (Chapters 6 and 7), with the preliminary studies needed to support those approaches (Chapters 4 and 5). The results presented here improve our understanding of road development and supplement previous efforts to model future road network development (e.g. Arima et al 2008, Jiang 2007).

8.2. Have the objectives of this thesis been addressed?

An attempt to link road networks with species richness levels was made in Chapter 3 to determine if there was (1) relationship between roadless volume and biodiversity, and (2) what predictions of species richness would look like extrapolated from this relationship. It was found that the metric, roadless volume, had a strong relationship with forest bird richness, explaining 72% of the observed variance. Higher species richness was found at higher roadless volumes, this is perhaps unsurprising as a high roadless volume indicates an area is less infiltrated by roads and thus is likely to experience less disturbance. Indeed roadless volume performed better as a predictor than percent forest cover, which is a commonly used richness indicator (Steffan-Dewenter 2002, Homen *et al.* 2004, Radford *et al.* 2005). This may be because roadless volume offers a metric that acts as a proxy for more cryptic forms of threats to biodiversity than just habitat loss. For example, habitat fragmentation which disrupts movement, forest degradation, fire and hunting (Peres *et al.* 2006, Lees & Peres 2008). . This relationship was scale dependant, with water catchment area proving to be the most suitable scale at which to relate roadless volume with species richness. Roadless volume was also found to correlate with avian community composition and when extrapolated over the study region it was found that species richness was high along the banks of the river and in areas that are as of yet undeveloped. Further, the areas projected to have the highest change in species richness lay along the frontiers of road development (which are also likely to be the frontiers of habitat change). The study region in this chapter (3) was fairly small, and it would be necessary to expand the analyses presented to other geographical locations in the Amazon. It would also be beneficial to conduct such analyses on taxa other than forest birds, small mammals or herpeto fauna for instance, to investigate if the observed relationship between roadless volume and species richness follows a similar relationship in other taxa or if the observed relationship is unique to forest birds. The initial literature review (Chapter 2) found no fewer than eight broad categories of road induced ecological effects, of which altered biodiversity is just one sub-category. It would be interesting to extend this analysis to assess if roadless volume has a relationship with any other ecological effects, such as animal range size, community succession or pollution.

The literature search (Chapter 2) revealed that unofficial road building, with no central management, is common place in the Brazilian Amazon, and much of the initial road network is laid by extractive industries (Laurance *et al.* 2009, Caldas *et al*. 2010). Logging is one of the most prevalent forms of extractive industry in the region and is likely to continue with the release of government logging concessions (SFB 2013). Upto two thirds of roads have been found to have been laid by loggers in frontier areas of Brazil (Uhl *et al.* 1991). Subsequent roads or other developments spring from this initial network, thus the decision to focus on modelling logging roads in Chapter 7 was made to address the thesis aim of generating amazon-wide road models. To this end the spatial pattern of economically valuable timber was explored (Chapter 4), with the aim of determining if there was a pattern to the distribution of timber value across the Amazon forest (which would influence where logging roads are likely to be built) and if this pattern could be related to ecological processes. This investigation revealed clear spatial patterns of economic value that were explained by known ecological processes relating to soil and climate. Findings showed that the most valuable timber tends to occur in the north east region of the Amazon; an area of poor soil quality, that is subject to occasional drought, making it particularly suited to slow growing hard woods that command high market values (Malhi *et al.* 2002, Baker *et al.* 2004 ter Steege *et al.* 2006). Although there was evidence of some species specific variation in this pattern, in the genera Mora and Dicorynia for example, the general pattern is believed to be robust.

The value patterns found in Chapter 4 are reasonable and useful for determining large scale patterns of road development, more local scale development may need finer resolution timber data. One complication that arises is that the specific timber species that are considered valuable changes through time, especially with the introduction or abolition of laws and policy on which species may or may not be harvested and traded. One way to overcome this would be to have a dynamic revenue layer, where the economic value of the genera may be updated and the specific genera considered by the krigging analysis may also be altered, such that only trees of current or potential interest are considered. Beyond the issues of market forces and data age however, the analysis of potential timber revenue available from the forest for use in road modelling is sound.

The objectives of Chapter 5 were to; (1) establish what the temporal dynamics of road density looked like, (2) determine over what time scales the phases of road development occur, and (3) to relate the observed patterns to anthropogenic and economic phenomena. Observations of past road development via a space for time substitution revealed that road development is density dependent following a 'boom and bust' dynamic through time that can be best described by a logistic curve. This logistic dynamic can be broken into three phases, the first (lag phase) lasted an average of 15 years, this was followed by a boom phase that lasted an average of 39 years, finally followed by a bust phase where development slows and eventually stops. It was also found that it takes on average 75 years for a road network to develop to the maximum observed density. The logistic nature found, suggests a 'boom and bust' pattern of development similar to commodity trends observed within the region, and linked with human wellbeing metrics. (Godfrey 1990, Macedo & Anderson 1993, Clough *et al.* 2009, Rodrigues *et al.* 2009). . A very small number of municipalities with very low initial road densities had road network growth rates that were much higher than the general pattern observed in the remaining municipalities. These apparently sporadic events of rapid road construction at particular places and times suggests that the general pattern of road network development can be accelerated under exceptional circumstances. The number of these exceptions was found to be very low, indicating that the general logistic nature of development is accurate. While we believe that the patterns found are realistic, additional past spatial road data would be useful for confirming the trajectory of road development found in Chapter 5 to determine if the observed range of trajectories is supported by empirical temporal data.

Confirmation of the logistic nature of road development led to the exploration of a spatial road development model inspired by models from the field of population ecology (Chapter 6). Here we sought to determine (1) if road development occurs in a directional manner (2) if models derived from population ecology can predict road density development and (3) if the processes governing road dynamics are intrinsic only, or if neighbourhood effects also play a role. Anisotropy analysis results confirm qualitative assumptions about road development gleaned from the literature (Fearnside 2008, Perz *et al.* 2008), indicating that at large scales road development is indeed directional with a tendency to move towards the deforestation frontier of the Amazon. It was found that although population based models can describe the large scale dynamics of roads, there is still plenty of mileage in improving these models as they were found to have low coefficients of correlation (~0.2). However, the time taken for road networks to develop in these density based models (approximately 60 years) is similar to the time estimated in Chapter 5. Indicating that these models are making valid, if not perfect, predictions. Finally, this modelling approach found that neighbourhood effects in addition to intrinsic density dependence does influence the future of road development. These models would benefit from the addition of co-factors of development other than neighbourhood to increase the amount of variance explained by the models. For example, friction data that would modulate the rate of development in different regions.A short fall of this modelling approach is that it generates large scale density predictions, while these predictions are useful, most land use and biodiversity change models rely on road maps

(Messina & Walsh 2001, Soares-Filho *et al.* 2002, Alkemade *et al.* 2009, Chapter 3) that this type of modelling cannot produce.

The process based model (Chapter 7) addresses this issue to an extent by imposing attractor (revenue) and friction (cost) layers, and sought to answer the question of whether a process based approach (based on costs and revenue) could accurately predict road networks. This modelling approach allows the incorporation of the economic ‘behaviour’ of loggers who build the roads. It is clear from the results presented in Chapter 7 that our model has not accurately replicated the patterns of road network expansion in the Brazilian Amazon, yet despite the poor overall predictions we believe the model approach has sound rationale and represents several advantages over the current suite of available models. Firstly we estimate uncertainty, which other road models omit, but which is clearly an important aspect of any prediction. Further, these model predictions cover a larger spatial extent and are more rigorously validated than existing validated models. A key benefit to this modelling approach is that any factor that determines the attractiveness of land for a road destination can be incorporated into the revenue layer, not just timber value. For example, agriculture, mining, and oil and gas extraction all provide a strong impetus for the development of roads in the Amazon (Laurance *et al.* 2004, Fearnside 2007, Finer *et al.* 2008), and it would be possible to include maps of known planned roads through official development plans such as IIRSA (Kileen 2005) and Pelt-Para (PELT-Pará 2012). An objective weighting method to select appropriate inputs and weights for the friction map needs to be developed and implemented to improve the cost aspect of the process model. One further development would be to include the use of actual financial costs of building a road. If we could obtain data on the real costs of road building more detailed profit analyses of potential roads could be conducted.

While this thesis has been largely successful in addressing the aims set out at the start of the study, it has led to further questions such as: How do we move forward to improve our predictive ability? Would this improved predictive power lead to a general understanding of how to predict road development in different locations? and How do we use these predictions to investigate various ecological effects at different scales?

8.3. Issues to resolve in order to move forward

In conducting the work presented, several difficulties arose, chief among them is the issue of data availability; there is a distinct lack of historical spatial data on road development in the Amazon. This data is imperative for model calibration, parameter fitting and validation. Throughout the work presented here 27 Landsat scale roadmaps covering an annual period of nine years (2000-2008) and two Amazon wide maps covering two time points (2004/2007) were used. All of these maps were manually digitised from Landsat TM satellite images; this process is time consuming and labour intensive, with each year/Landsat combination taking approximately 10-15 hours to complete depending on the road density. Automated approaches to extracting road location information from images exist but it is generally argued that these approaches are far less accurate than manual digitisation (Mena 2003, Brandao & Souza 2006, Li & Briggs 2009, Movaghati *et al.* 2010). An obvious advancement would be to improve automated methods to increase the amount of data available upon which to base models. This would increase the temporal range of the data to test if road development truly does follow a logistic pattern through time, as shown by the presented space-for-time substitution in Chapter 5. It would also allow the models presented in Chapters 6 and 7 to be validated from various historical start points, to see if start point has an effect on predictive power or if model accuracy diminishes through time at a similar rate regardless of start point (i.e. does development behave differently in different eras, for instance the 1980's vs. 1990's). Further, an increase in road data would lead to more robust model calibration and would allow the possibility of parameter inference that is unlikely to be successful with the limited data currently available.

In addition to spatial road data for calibration and validation of the predictive models, better data on the other model inputs are needed. Considering the process based model (Chapter 7) the data used to estimate timber revenue was collected in the 1970s as part of the RADAMBRAZIL survey. As of 2011 the Brazilian SFB (forestry service of Brazil) in collaboration with the FAO (Food and Agriculture Organisation of the United Nations) plans to conduct a massive nationwide forest monitoring system beginning with a field inventory of trees within the region on a 20km grid, where the intersection of each grid square is a sample point (FAO 2013). This project is also supported by GEF (global environment facility) who are investing nearly nine million dollars, it is hoped that this inventory will be completed by 2015. This inventory is not only of use to update the revenue map used in the model but also to investigate the potential effects of road development, because the inventory is also recording information on carbon stocks, biodiversity and the social importance of forests to local communities (IFN National Forest Inventory 2013).

It is not only road and model input data availability that needs to be improved, in order to link roads to ecological effects better monitoring of the environment before and after roads are laid needs to be carried out. Biodiversity is possibly the 'best effect' to study as there is already wide ranging interest in factors that affect and alter biodiversity levels, thus we chose to relate roadless volume to species richness in Chapter 3. As with spatial road data, temporal data of species richness is necessary to validate predicted changes in species richness with roadless volume with observed changes to richness. While biodiversity changes may be an obvious choice of ecological effects to study, other road impacts both ecological and social could be considered with road modelling, for example the spread of a pathogen.

While it is freely admitted that data availability and accessibility is poor it is not obvious on how this situation can be easily remedied practically. This is because data is expensive and often difficult to produce and even when the data does exist it generally does so on someone's hard drive or in data repository, where data keys are all but unintelligible except to the data creators. This issue with open, intelligible and easily accessible data is not a unique problem to the fields of development, ecology or to the Amazon. Appendix D (a review of land use land cover change models) highlights and discusses this issue in detail. It is sufficient here to re-iterate that it is a barrier to not only model development and implementation, but also to decision makers who are potentially basing decisions on less data than necessary.

In addition to improving models with better and larger quantities of data, an important aspect of modelling to consider is uncertainty. There are many definitions of uncertainty which vary between authors and fields, reflecting underlying differences in approach and/or philosophy

(Regan *et al.* 2002, Walker *et al*. 2003, Refsgaard *et al*. 2007). Examples of uncertainty include (but are certainly not limited to), measurement error, systematic error, inherent randomness, model uncertainty (defined as errors in our representation of the real world in a model), natural variation and linguistic uncertainty (where the use of language without fixed definitions confuses matters and results in uncertainty) (Regan *et al*. 2002), model uncertainty (defined as uncertainty associated with either or both, model structure and model technical uncertainty), input uncertainty, parameter uncertainty, model outcome uncertainty (Walker *et al.* 2003). Clearly, the term 'uncertainty' is broad, multi-faceted, with many meanings (Montanari 2007) and suffers from linguistic uncertainty as defined by Regan *et al*. (2002). Here uncertainty describes the confidence we have in our predictions (model outcome uncertainty as defined by Walker *et al.* 2003). For example, if we calculate a mean from sample data, we estimate confidence intervals that indicate whether our estimated mean is close to the 'true' mean, e.g. a mean of 60 with CIs of 61-62 is much better than a mean of 60 with CIs if 40-80. Any estimate made should take into account uncertainty and model predictions are no different. Uncertainty around a mean arises from sampling or measurement discrepancies, while uncertainty in a model prediction has three sources; (1) data uncertainty, (2) stochastic model uncertainty, and (3) structural model uncertainty.

Data uncertainty refers to the degree to which input data varies; nearly all model input data are subsets (e.g. a sample) or derivations (e.g. an average) of all possible data. This type of uncertainty includes aspects such as, measurement error, systematic error, natural variation (Regan *et al*. 2002), input choice and parameter uncertainty (Walker *et al.* 2003). Thus the exact data used for modelling will vary depending on how the data were collected, processed or selected. This variation could lead to potential shifts in model predictions, for example if we were to model roads using input data on a 250m grid vs. a 1km grid, the overall road network may look similar but the exact route taken by the roads will vary, or if we used a revenue map with the genera Havea replacing Mora we could see differing specific destination locations because the interpolation of revenue would be slightly altered. This type of uncertainty can be quantifiyed using sensitivity analyses, where the data are deliberately altered for each model iteration and the resulting predictions are compared. If there are large differences in the predictions then there is said to be a lot of data uncertainty. There is little we can do to reduce this type of uncertainty, except trying to improve the amount, coverage and resolution of the input data.

Stochastic model uncertainty comes not from the data but from stochastic processes within the model itself. For example, in the road process model (Chapter 7) each iteration allows a different set of road destinations to be chosen, this alters the predictions and introduces uncertainty. This uncertainty can be quantified by running the model multiple times (with the same input data) and estimating how different the predictions are from each other or the probability of any given prediction occurring. In other words, by conducting sensitivity analysis, this type of exrceise is often used by modellers to quantify prediction uncertainty arising from internal processes of a model (Klepper 1997, Minunno *et al*. 2013, Wang *et al*. 2013).

Structural model uncertainty is related to how well the model structure, process and function, represent the real world (termed 'model uncertanty' by Regan *et al.* 2002). For example, a model that captured and described all processes in a system accurately and placed those processes in the correct order with the correct interactions is likely to have very low structural uncertainty, as the structure of the model closely matches that of the real world thus the predictions are likely to match observations. This type of uncertainty arises from how the model 'works' rather than from any stocastisity within the model or variation in input data. This type of uncertainty could potentially be quantified by validating predictions with observations, while taking into account data and model uncertainty. This however is exceedingly difficult to do, indeed Regan *et al.* (2002) suggest this type of uncertainty 'is notoriously difficult to quantify and impossible to eliminate', thus, validation of predictions against observations generally describes overall 'prediction uncertainty' (or 'model outcome uncertainty' as it is termed by Walker *et al.* 2013).

Despite the fact it would be difficult to disentangle the three sources of prediction uncertainty, a measure of uncertainty should be reported on any predictions made. The concept of prediction uncertainty (regardless of whether it arises from data, stochastic model or structural model uncertainties) is very important not only for assessing the performance of a model (e.g. judging if it is a good or poor model) but also for the utility of a model in a wider sense (e.g. decision making). A predictive model is useful in two key ways; (1) for scenario analyses, where various possibilities are explored and assessed but little weight is given to individual predictions, (2) as a predictive tool, where an assumption that the predictions are 'true' is made and credence is given to individual predictions. If uncertainty is not considered, there can be serious implications if decisions are based on predictions. A key example of how important uncertainty is for model utility in decision making and the potential disastrous consequences of not taking it into consideration, is the collapse of the Canadian cod fisheries in the 1980/90's. The uncertainty of fish stock assessments was downplayed and decisions on catch allowance were made assuming that models were using 'certain' data. This was not the case and the decisions made by fishery managers caused the collapse of no less than seven fisheries. It is argued that there were other contributing factors to this collapse, but the role of model uncertainty cannot be ignored. Had this uncertainty been considered perhaps a 'precautionary approach' may have been taken, with lower quotas (Walters & Maguire 1996). The importance of uncertainty applies across all models, including road and LULC models, upon which any policy or decision could be based. While it is not the role of a scientist to influence policy or decisions, it is their responsibility to help decision makers understand the uncertainties arising from their models and ensure that they are taken into account in assessments. Which is why such rigorous validation methods were applied to the road models present in this thesis.

8.4. Future utility of models presented

Ultimately this work was done to assess the ecological implications of roads, on ecological processes, function and services; What effects do roads have? How do we connect roads and impacts? and, What does development mean for ecology in the Amazon and more widely? In order to answer these questions it would be ideal to get the models presented to a state where they can be used in conjunction with other tools and models (especially LULC models) to run scenarios and aid decisions in terms of policy, and also perhaps in terms of industry decisions. With the rise in green, 'eco', and ethical companies, there is a significant effort to factor the environment into business decisions and these companies have an ability to make a real difference (Bansal & Roth 2000, Beveridge & Guy 2005, Chen *et al*. 2006, Dangelico & Pujari 2010). Using models to help plan decisions and assess impacts could help determine what is a better/worse in terms of road network design given all that is wanted to be achieved over time, for instance; access to timber, a useful infrastructure foundation network for subsequent users, a network that has lower environmental impacts, and a network that offers value for money. Indeed Laurance & Balmford (2013) recently argued for a global road zoning exercise to aid transport planning with the environment in mind.

In the work presented, we have reasonably predicted road development. The predictions are not completely accurate, but they are reasonable and they are useful. The nature of road development in the Amazon is inherently highly stochastic, with small decisions being made by loggers having an impact on where and how the network will develop. For example, the choice to move north vs. north west seeking timber can dramatically alter the direction that the roads develop in. Thus we expect individual iterations of models not to match the observed network. Rather, what we can expect, is to capture the general patterns of development and apply probabilities to our estimations, which, is essentially what we have done. Further, we can expect that at larger scales our predictions reflect more the nature of development, i.e. while we may not be able to predict the exact route a road is going to take, we can show that the roads are moving north-westerly with a given amount of road added in each time step. Even at small scales, individual roads, if we have the correct destinations, the process model can predict roads with good accuracy (improvements to destination selection are immensely important here).

It is likely that the models presented here could be applied to road development in other tropical regions where the logging industry is prevalent and at the forefront of development. The models are unlikely to be applicable to forestry or road development in developed regions, for instance, Northern Europe or Canada, despite the fact that forestry is a large industry here. This is mainly because of differing management scenarios, in tropical regions decisions are often made by on the ground loggers, while in developed regions decisions on where to build roads are carefully assessed and often undertaken by management personnel. Natural primary forest logging as seen in the tropics is very different from plantation/secondary/re-planted logging as seen in more developed regions (Gray 2002), as such the processes driving road network development is very different. A further nuance is the difference between individuals, small and large logging operations. In developed regions logging is controlled largely by large organisations, in Brazil however there is a lot of variation in who is carrying out the logging and road building.

While spatially the process model could be applied to other tropical regions, temporally it is specific to the start of the road development process in frontier regions. In the early stages, logging is the driving force of network development, later as colonisers and agriculture moves into an area we see different dynamics take over. Specifically, the dendritic nature of logging roads is replaced by gridded structure road networks that branch off from the initial logging roads. Thus the process based model (Chapter 7) is only applicable in the first stage of development after which another model/module would need to implemented to capture subsequent development. The process model could probably only make location predictions on the time scales of logging concessions, on average 40-50 years. The density model (Chapter 6) however, is not subject to this caveat because it is parameterised based on all roads and only considers how the overall network density changes through time and space (not just logging roads). Another issue that affects the models' ability to predict networks into the future is the effects of legislation, local and global markets and economies that are not counted in the current models. It is possible to generalise the models (Chapter 6 and 7) across tropical regions, assuming that they are re-calibrated and validated for each new region. Temporally however the process model is only relevant in the initial stages of development on logging frontiers, while the density model is applicable to any stage of development.

The road models presented are important not just for understanding and predicting road network development but also for incorporating into land use land cover change models. It is widely accepted that roads are a key predictor of deforestation (Fearnside 2008; Geist & Lambin 2002; Perz *et al.* 2007; Perz *et al.* 2008), that roads are difficult to predict (Barlow *et al.* 2011) and that current predictions are inadequate (Rosa et al. 2014). The models presented here could help improve land use change models by improving a key input to said models. Even if exact or long term predictions of the network development cannot be achieved, the predictions from these models go a way to improve the situation. Scenario analyses can still be conducted and useful insight can be gained. Tradeoffs of roads developing in different regions can be compared, and the probability of development can help indicate determine where protection or policing should be concentrated.

Roads in the Amazon, currently, develop initially for logging (and other extractive industries), but are later used for secondary or tertiary reasons long after the initial use has ceased, i.e. logging roads are used for a number of years, but are then used by the agricultural industry and the general population long after logging has stopped. But, the roads built for the primary purpose may not be optimal for subsequent users. One way in which road location based models can be used is to look at the values of different uses over time and attempt to influence the development of roads not just to reduce environmental impacts but also to influence development for a more optimal road network over the long term.

8.5. Conclusion

We have developed a frame work for showing the uncertainty around road maps (Chapter 7) which previously did not exist. Prior to this, spatial predictions of road networks have been single output maps, where our results lend themselves to probability maps and example single maps. We have shown it is possible to link biodiversity levels with road networks via road metrics other than distance to nearest road or road density (Chapter 3), finding a high correlation between roads and species richness. We have established the density dependant nature of road development and determined via two independent analyses (Chapter 5 and 6) that it would take approximately 70 years for a network to develop to maximum density from initial development in a frontier region.

Results of this thesis relate to Amazon road development and has led to new insights about how difficult it is to predict development, it has also raised questions about how we currently predict roads, if this is the best approach and if not, what are the alternatives.

The Amazon is the largest area of natural forest habitat it is important to plan its use well. Decisions of how to use the forest are being made at all scales, from global (with the implementation of REDD+ schemes), to regional (Brazilian government allowing new concessions, and regulating timber), to local government, and individual farms (how much of the land is being left forested and how much planted?) or individuals (using the forest in a subsistence way). Models such as those proposed here in this thesis are aimed at the larger scale decision makers, from local government or firms to global regulators.

Roads are one of the first stages in the human/ecosystem interface. Many people talk about human interaction with the environment as if humans are somehow outside of the system; the truth is we are very much a part of the system and decisions we make in altering our environment has long lasting impacts, many of which we are unable to predict. It is hoped that the models presented here will be improved and play a role in aiding those decisions.

References

Reference list

Adams, L. W. & Geis, A. D. (1983) Effects of roads on small mammals, Journal of Applied Ecology, 20, 403-415.

Ahmed, S. E., Souza, C. M., Riberio, J. & Ewers, R. M. (2013) Temporal patterns of road network development in the Brazilian Amazon, Regional Environmental Change, InPress: DOI 10.1007/s10113-012-0397-z

Alkemade, R., van Oorschot, M., Miles, L., Nellemann, C., Bakkenes, M. & ten Brink, B. (2009) GLOBIO3: a framework to investigate options for reducing global terrestrial biodiversity loss. Ecosystems, 12, 374-390.

Alves, D. S. (2002) Space-time dynamics of deforestation in Brazilian Amazon, International Journal of Remote Sensing, 23,2903-2908.

Andersen, L. E & Reis, E. J. (1997) Deforestation, development, and government policy in the Brazilian Amazon: an econometric analysis, IPEA (Institudo de Pesquisa Economica

Angold, P. G. (1997) The impact of a road upon adjacent heathland vegetation: effects on plant species composition, Journal of Applied Ecology, 34, 409-417.

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

Arima, E. Y., Walker, R. T., Perz, S. G. & Caldas, M. (2005) Loggers and forest fragmentation: behavioural models of road building in the Amazon Basin, Annals of the Association of American Geographers, 95, 525-541.

Arima, E. Y., Walker, R. T., Sales, M., Souza Jr, C. & Perz, S. G. (2008) The fragmentation of space in the Amazon basin: emergent road networks, Photogrammetric Engineering & Remote Sensing, 74, 699-709.

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

Asner, G. P., Keller, M. & Silvas, J. N. M. (2004a) Spatial and temporal dynamics of forest canopy gaps following selective logging in the Eastern Amazon, Global Change Biology, 10, 765-783.

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

Asner, G. P., Nepstad, D., Cardinot, G. & Ray, D. (2004b) Drought stress and carbon uptake in an Amazon forest measured with spaceborne imaging spectroscopy, PNAS, 101, 6039-6044.

Asner, G.P., Broadbent, E.N., Oliveira, P.J.C., Keller, M., Knapp, D.E. & Silva, J.N. (2006) Condition and fate of logged forests in the Brazilian Amazon. Proceedings of the National Academy of Sciences 103, 12947–12950

Bain, M. B., Finn, J. T &Booke, H. E. (1988) Streamflow regulation and fish community structure, Ecology, 69, 382-392.

Bairlein, F., Norris, D. R., Nagel, R., Bulte, M., Voigt, C. C., Fox, J. W., Hussell, D. J. T. &Schmaljohann (2012) Cross- hemisphere migration of a 25g songbird, Biology Letters, 8, 505-507.

Baker, T. R., Phillips, O. L., Laurance, W. F., Pitman, N. C. A., Almeida, S., *et al.* (2009) Do species traits determine patterns of wood production in Amazonian forests?, Biogeosciences, 6, 297-307.

Baker, T. R., Phillips, O. L., Malhi, Y., Almeida, S., Arroyo, L., *et al.* (2004) Variation in wood density determines spatial patterns in Amazonian forest biomass, Global Change Biology, 10, 545-562.

Baldi, A. & Kisbenedek, T. (1999) Species- specific distribution of reed-nesting passerine birds across reed-bed edges: effects of spatial scale and edge type, Acta Zoologica Academiae Scientiarum Hungaricae, 45, 97-114.

Bansal, P. & Roth, K. (2000) Why companies go green: a model of ecological responsiveness, The Acadamy of Management Journal, 43, 717-736.

Barlow, J., Ewers, R. M., Anderson, L., Aragao, L. E. O. C., Baker, T. R., Boyd, E., Feldpausch, T. R., Gloor, E., Hall, A., Malhi, Y., Milliken, W., Mulligan, M., Parry, L., Pennington, T., Peres, C. A., Phillips, O. L., Roman-Cuesta, R. M., Tobias, J. A. & Gardner, T. A. (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.

Barreto, P., Amaral, P., Vidal, E. & Uhl, C. (1998) Costs and benefits of forest management for timber production in eastern Amazonia, Forest Ecology and Management, 108, 9-26.

Barthelmess, E. L. & Brooks, M. S. (2010) The influence of body size and diet on road-kill trends in mammals, Biodiversity conservation, 19, 1611-1629.

Bazzaz, F. A. & Garbutt, K. (1988) The response of annuals in competitive neighbourhoods: effects of elevated CO2, Ecological Society of America, 69, 937-946.

Beaudry, F., deMaynadier, P. G. & Hunter, M. L. Jr. (2008) Identifying road mortality treat at multiple spatial scales for semi- aquatic turtles, Biological Conservation, 141, 2550-2563.

Bee, M. A., & Swanson, E. M. (2007) Auditory masking of anuran advertisement calls by road traffic noise, Animal Behaviour, 74, 1765-1776.

Begon, M., Harper, J. L. & Townsend, C. R. (1996) Ecology, Third Edition, Blackwell Publishing Company, Oxford.

Belloc, H. (1923) The road, The British Reinforced Concrete Engineering Co LTD., Manchester.

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

Beyer, H. L. 2004.Hawth's Analysis Tools for ArcGIS.Available at: http://www.spatialecology.com/htools

Beveridge, R. & Guy, S. (2005) The rise of the eco-preneur and the messy world of environmental innovation, International Journal of Justice and Sustainability, 10, 665-676.

Bhattacharya, M., Primack, R. B. & Gerwein, J. (2003) Are roads and railroads barriers to bumblebee movement in a temperate suburban conservation area?, Biological Conservation, 37-45.

Bingal, K. L., Ashmore, M. R., Headley, A. D., Stewart, K. & Weigert, K. (2007) Ecological impacts of air pollution from road transport on local vegetation, Applied Geochemistry, 22, 1265-1271.

Blom, A., van Zalinge, R., Heitkonig, I. M. A. & Prins, H. H. T. (2005) Factors influencing the distribution of large mammals within a protected central African forest, Oryx, 39, 381-388.

Bonan, G. B. (2008) Foresta and climate change, forcings, feedbacks and the climate benefits of forests, Science, 320, 1444-1449.

Brack, D. (2003) Illegal logging and the illegal trade in forest and timber products, International Forestry Review, 5, 195-198.

Bradshaw, C. J. A., Sodhi, N. S., Peh, K. S. H. & Brook, B. W. (2007) Global evidence that deforestation amplifies flood rish and severity in t he developing world, Global Change Biology, 13, 2379-2395.

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

Brandão, Jr, A. O., Souza Jr, C. M., Ribeiro, J. G. F. & Sales, M. H. R. (2007) Desmatamento e estradasnao-oficiais da Amazonia. Anais XIII SimpósioBrasileiro de SensoriamentoRemoto, 2007, 2357-2364.

Brisson, J., de Blois, S. & Lavoie, C. (2010) Roadside as invasion pathway for Common Reed (Phragmites australis), Weed Science Society of America, 3, 506-514.

Broadbent, E. N., Asner, G. P., Keller, M., Knapp, D. E., Oliveira, P. J. C. & Silva, J. N. (2008) Forest fragmentation and edge effects from deforestation and selective logging in the Brazilian Amazon, Biological Conservation, 141, 1745-1757.

Brown, G. P., Phillips, B. L., Webb, J. K. & Shine, R. (2006) Toad on the road: use of roads as dispersal corridors by cane toads (Bufo marinus) at an invasion front in tropical Australia, Biological Conservation, 133, 88-94.

Brumm, H. (2004) The impact of environmental noise on song amplitude in a territorial bird, Journal of Animal Ecology, 73, 434-440.

Brumm, H., Schmidt, R. & Schrader, L. (2009) Noise-dependant vocal plasticity in domestic fowl, Animal Behaviour, 78, 741-746.

Brumm, H., Voss, K., Kollmer, I. & Todt, D. (2004) Acoustic communication in noise: regulation of call characteristics in a new world monkey, The Journal of Experimental Biology, 207, 443-448.

Caceres, N. C. (2011) Biological characteristics influence mammal road kill in an Atlantic Forest-Cerrado interface in south-western Brazil, Italian Journal of Zoology, 78, 3, 379-389.

Cade, B. S., Terrell, J. W. & Schroeder, R. L. (1999) Estimating effects of limiting factors with regression quantiles, Ecology, 80, 311-323.

Caldas, M. M., Simmons, C., Walker, R., Perz, S., Aldrich, S., Pereira, R., Leite, F. & Arima, E. (2010) Settlement formation and land cover use change: A case study in the Brazilian Amaon, Journal of Latin Amerian Geography, 9, 125-144.

Calderon, C. &Serven, L. (2004) The effects of infrastructure development on growth and income distribution, Central Bank of Chile Working papers, number 270, Available from http://www.bcentral.cl/estudios/documentos-trabajo/pdf/dtbc270.pdf (Accessed 12/12/2012).

Cameron, E. K. & Bayne, E. M. (2009) Road age and its importance in earthworm invasion of northern boreal forests, Journal of Applied Ecology, 46, 28-36.

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

Carr, L . W. & Fahrig, L. (2001) Effect of road traffic on two amphibian species of differing vagility, Conservation Biology, 15, 1071-1078.

Carvalho, G. O., Nepstead, D., McGrath, D., del Carmen Vera Diaz, M. *et al.* (2002) Frontier expansion in the Amazon, Environment, 44, 34-45.

Caughley, G. (1994) Directions in conservation biology, Journal of Animal Ecology, 63, 215-244.

Chave J, Muller-Landau HC, Baker TR, Easdale TA, ter Steege H, Webb CO (2006) Regional and phylogenetic variation of wood density across 2456 neotropical tree species, Ecological applications, 16, 2356-2367.

Chen, Y., Lai, S. & Wen, C. (2006) The influence of green innovation performance on corporate advantage in Taiwan, Journal of Business Ethics, 67, 331-339.

Chomitz, K. M. & Gray, D. A. (1996) Roads, land-use, and deforestation: a spatial model applied to Belize, World Bank Economic Review, 10, 487-512.

Chruszcz, B., Clevenger, A. P., Gunson, K. E. & Gibeau, M. L. (2003) Relationships amoung grizzly bears, highways and habitat in the Banff-Bow Valley, Alberta, Canada, Canadian Journal of Zoology, 81, 1378-1391.

Cinzano, P., Falchi, F. And Elvidge, C. D. (2001) The First World Atlas of the Artificial Night Sky Brightness, Mothly Notices of the Royal Astronomical Society, 328, 689-707.

Clark Labs (2007) The Land Change Modeler for Ecological Sustainability. IDRISI focus paper, Available from: http://clarklabs.org/applications/upload/Land-Change-Modeler-IDRISI-Focus-Paper.pdf Accessed June 2013.

Clark, W. D. & Karr, J. R. (1979) Effects of highways on red-winged blackbird and horned lark populations, The Wilson Bulletin, 91, 143-145.

Clevenger, A. P., Chruszez, B. & Gunson, K. E. (2003) Spatial Factors Influencing Small Vertebrate Fauna Road-Kill Aggregations, Biological Conservation, 109, 15-26.

Clough, Y., Faust, H. &Tscharntke, T. (2009) Cacao boom and bust, sustainability of agroforests and opportunities for biodiversity conservation, Conservation Letters, 2, 197-205.

Cochrane, M. A. & Laurance, W. F. (2002) Fire as a large-scale edge effect in Amazonian forests, Journal of Tropical Ecology, 18, 311-325.

Coffin, A. W. (2007) From Roadkill to Road Ecology: A Review of Ecological Effects of roads, Journal of Transport Geography, 15, 396-406.

Colchero, F., Conde, D. A., Manterola, C., Chavez, C., Rivera, A. & Ceballos, G. (2011) Jaguars on the move: modeling movement to mitigate fragmentation from road expansion in the Mayan forest, Animal Conservation, 14, 158-166.

Colwell, R. K. 2009. EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.2. User's Guide and application published at: http://purl.oclc.org/estimates.

Costa, M. H., Botta, A. & Cardille, J. A. (2003) Effects of large scale changes in land cover on the discharge of the Tocantins river, South-eastern Amazonia, Journal of Hydrology, 283, 206-217.

Crawley, M.J. 2008 The R Book Chichester: Wiley.

Cynx, J., Lewis, R., Tavel, B. & Tse, H. (1998) Amplitude regulation of vocalizations in noise by a songbird, Taeniopygia guttata, Animal Behaviour, 56, 107-113.

Da Silva, J. M. C., Rylands, A. B. & Da Fonseca, G. A. B. (2005) The fate of the Amazonian areas of endemism, Conservation Biology, 19, 689-694.

Dale, V. H., O'Neill, R. V., Southworth, F. & Pedlowski, M. (1994) Modelling Effects of land management in the Brazilian settlement of Rodonia, Conservation Biology, 8, 196-206.

Dangelico, R. M. & Pujari, D. (2010) Mainstreaming green product innovation: why and how large companies integrate environmental sustainability, Journal of Business Ethics, 95, 471-486.

Dauvergne, P. & Lister, J. (2012) Timber, Polity Press, Cambridge.

Davidson, E. A., Savage, K. E., Bettez, N. D., Marino, R. & Howarth, R. W. (2010) Nitrogen in runoff from residential roads in a coastal area, Water Air Soil Pollution, 210, 3-13.

de Barros Ferraz, S., Vettorazzi, C. A., Theobald, D. M. & Bellester, M. V. R. (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.

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

De Oliveira Filho, F. J. B. & Metzger, J. P. (2006) Thresholds in landscape structure for three common deforestation patterns in the Brazilin Amazon,Landscape Ecology, 21, 1061-1073.

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

Dean, W. R. J & Milton, S. J. (2003) The importance of roads and road verges for raptors and crows in the succulent and Nama-Karoo, South Africa, Ostritch, 74, 181-188.

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

Delgado, J, D., Arroyo, N. L., Arevalo, J. R. & Fernandez-Palacios, J. M. (2007) Edge effects of roads on temperature, light, canopy cover and canopy height in laurel and pine forests (Tenerife, Canary Islands), Landscape and Urban Planning, 81, 328-340.

Develey, P. F. & Stouffer, P. C. (2001) Effects of roads on movements by understory birds in mixed species flocks in central Amazonin Brazil, Conservation Biology, 15, 1416-1422.

Dirzo, R. & Raven, P. H. (2003) Global state of biodiversity and loss, Annual Review of Environment and Resources, 28, 137-167.

Dobson, A., Borner, M., Sinclair, T. & 24 others (2010) Road will ruin Serengeti, Nature, 467, 272-274.

Dong, S. K., Cui, B. S., Yang, Z. F., Liu, S. L., Liu, J., Ding, Z. K., Zhu, J. J., Yao, W. K. & Wei, G. L. (2008) The role of road disturbance in the dispersal and spread of Ageratina adenophora along the Dian-Myanmar international road, Weed Research, 48, 282-288.

Donovan, T. M., Jones, P.W., Annand, E. M. & Thompson, F. R. (1997) Variation in local-scale effects: mechanisms and landscape context, Ecology, 78, 2064-2075.

Dyer, S., O'Neill, J. P., Wasel, S. M. & Boutin, S. (2002) Quantifying barrier effects of roads and seismic lines on movements of female woodland caribou in northeastern Alberta, Canadian Journal of Zoology, 80, 839-845.

Ebenso, I. E. & Ologhobo, A. D. (2008) Effects of lead pollution from vehicular exhaust fumes against sentinel juvenile Achatina achatina, Bulletin of Environmental Contamination and Toxicology, 81, 513-515.

Eigenbrod, F., Hecnar, S. J. & Fahrig, L. (2008) The relative importance of road traffic and forest cover on anuran populations, Biological conservation, 141, 35-46.

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

Evans, T. P., Manire, A., De Castro, F., Brondizio, E., McCracken, S. (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. & Didham, R. K. (2006) Continuous response functions for quantifying the strength of edge effects, Journal of Applied Ecology, 43, 527-536.

Ewers, R. M., W. F. Laurance, & C. M. Souza Jr. (2008) Temporal fluctuations in Amazonian deforestation rates. Environmental Conservation 35, 303-310.

Fagan, W. F., Cantrell, R. S. & Cosner, C. (1999) How habitat edges change species interactions, The American Naturalist, 153, 165-182.

Fahrig, L. & Grez, A. A. (1996) Population spatial straucture, human-caused landscape changres and species survival, Revista Chilena de Historia Natural, 69, 5-13.

Fahrig, L. & Rytwinski, T. (2009) Effects on Animal Abundance: an Emperical Reveiw and Synthesis, Ecology and Society, 14, article 21.

Fajardo, I. (2001) Monitoring non-natural mortality in the barn owl (Tyto alba), as an indicator of land use and social awareness in spain, Biological Conservation, 97, 143-149.

FAO Food and Agriculture Organizaion (2013) http://www.fao.org/forestry/17847/en/bra/ Accessed June 2013

Farmer, A. M. (1993) The effects of dust on vegetation- a review, Environmental Pollution, 79, 63-75.

Fearnside, P. M. (1987) Deforestation and international economic development projects in Brazilian Amazonia, Conservation Biology, 1, 214-221.

Fearnside, P. M. (2005) Deforestation in Brazillian Amazonia, history, rates and consequences, Conservation Biology, 19, 680-688.

Fearnside, P. M. (2007) Brazils Cuiaba-Santarem (BR163) Highway: The Envieonmental Cost of Paving a Soybean Corridor Through the Amazon, Environmental Management, 39, 601-614.

Fearnside, P. M. (2008) The roles and movements of actors in the deforestation of Brazilian Amazonia, Ecology and Society, 13, 23-45.

Ferrari M.J., Grais R. F., Bharti N., Conlan A. J. K., Bjørnstad O. N., Wolfson L. J., Guerin P. G., Djibo A. & Grenfell B. T. (2008) The dynamics of measles in sub-Saharan Africa, Nature doi:10.1038/nature06509

Findley, S. C. & Bourdages, J. (2000) Response time of wetland biodiversity to road construction on adjacent lands, Conservation Biology, 14, 86-94.

Findley, S. C. & Houlahan, J. (1997) Anthropogenic correlates of species richness in Southeastern Ontario wetlands, Conservation Biology, 11, 1000-1009.

Finer, M., Jenkins, C. N., Pimm, S. L, Keane, B. & Ross, C. (2008) Oil and gas projects in the Western Amazon, threats to wilderness, biodiversity, and indigenous people, PLoSOne, 3, e2932.

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

Forbes, R. J. (1964) Ancient roads, second edition, Adolf M. Hakkert, Amsterdam.

Forman, R. T. T. & Alexander L. E. (1998) Roads and their major ecological effects, Annual Reviews of Ecological Systems, 29, 207-231.

Forman, R. T. T. & Deblinger, R. D. (2000) The ecological road- effect zone of a Massachusetts (USA) suburban highway, Conservation Biology, 14, 36-46.

Forman, R. T. T., (1998) Road Ecology: A solution for the giant embracing us, landscape Ecology, 13, 3-5.

Forman, R. T. T., Sperling, D., Bissonette, J. A., Clevenger, A. P., Cutshall, A. P., Dale, V. H., Fahrig, L., France, R., Goldman, C. R., Heanue, K., Jones, J. A., Swanson, F. J., Turrentine, T., Winter, T. C. (2003) Road Ecology Science and Solutions, Island Press, Washington.

Freeman, L. C. (1977) A set of measures of centrality based on betweenness, Sociometry, 40, 35-41.

Fuller, R. A., Warren, P.H. & Gaston, K. J. (2007) Daytime noise predicts nocturnal singing in urban robins, Biology Letters, 3, 368-370.

Gardner, T. A., Barlow, J., Cazdon, R., Ewers, R. M., Harvey, C. A., Peres, C. A. &Sodhi, N. J. (2009) Prospects for tropical forest biodiversity in a human-modifiyed world, Ecology Letters, 12, 561-582.

Gardner, T. A., Ferreira, J. F., Parry, L., Barlow, J., and 95 collaborators of the Sustainable Amazon Network. (2013). A social and ecological assessment of tropical land-uses at multiple scales: the Sustainable Amazon Network. Philosophical transactions of the Royal Society (Series B).

Gehlhausen, S. M., Schwartz, M. W & Augspurger, C. K. (2000) Vegetation and microclimatic edge effects in two mixed-mesophytic forest fragments, Plant Ecology, 147, 21-35

Geist, H., J. & Lambin, E. F. (2002) Proximate causes and underlying driving forces of tropical deforestation, BioScience, 52, 143-150.

Gelbard, J. L. & Belnap, J. (2003) Roads as conduits for exotic plant invasions in a semi-arid landscape, Conservation Biology, 17, 420-432.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004) Bayesian Data Analysis, Chapman and Hall.

Gentry, A. H. (1992) A synopsis of Bignoniaceae ethnobotany and economic botany, Annals of the Missouri Botanical Garden, 79, 53-64.

Gerwing, J. J., Johns, J. S., Vidal, E. (1996) Reducing waste during logging and log processing: forest conversion in eastern Amazonia, Unasylva, 187: article 3. Available online at: http://www.fao.org/docrep/w2149e/w2149e00.htm Accessed: 2012 Mar 18.

Ghazoul, J. & Sheil, D. (2010) Tropical Rainforest Ecology, Diversity and conservation, Oxford University Press, Oxford.

Gilks, W. R., Richardson., S. andSpiegelhalter, D. J. (1996) Markov Chain Monte Carlo in Practice, Chapman and Hall.

Giraudoux, P. R. (2011) pgirmess: Data analysis in ecology. R package version 1.5.1. http://CRAN.R-project.org/package=pgirmess Accessed: 2012 Apr 2.

Glista, D. J., DeVault, T. L. & Woody, J. A. (2009) A review of mitigation measures for reducing wildlife mortality on roadways, Landscape and Urban Planning, 91, 1-7.

Glover, D. R. and Simon, J. L. 1975. The Effect of Population Density on Infrastructure : The Case of Road Building. Economic Development and Cultural Change 23, 453-468.

Godfrey, B. J. (1990) Boom towns of the Amazon, American Geographical Society, 80, 103-117.

Goosem, M. (2007) Fragmentation impacts caused by roads through forests, Current Science, 93, 1587-1595.

Goosem, M. (2012) Mitigating the impacts of rainforest roads in Queensland's wet tropics: effective or are further evaluations and new mitigation strategies required?, Ecological Management & Restoration, 13, 254-258.

Gray, J. A. (2002) Forest concession policies and revenue systems: country experience and policy changes for sutainable tropical forestry, World Bank Technical paper, Number 552, Available from:

http://publications.worldbank.org/index.php?main_page=product_info&cPath=0&products_id=20637 Accessed: Dec 2013

Gregory, J. W. (1931) The story of the road, The university Press, Glasgow.

Haemig, P. D., Waldenstrom, J. & Olsen, B. (2008) Roadside ecology and epidemiology of tick-borne diseases, Scandinavian Journal of Infectious Diseases, 40, 853-858.

Halfwerk, W. & Slabbekoorn, H. (2009) A behavioural mechanism explaining noise-dependant freaquency use in urban bird song, Animal Behaviour, 78, 1301-1307.

Hall, A. (2008) Paying for environmental services: the case of the Brazilian Amazonia, Journal of International Development, 20, 965-981.

Haskell, D. G. (2000) Effects of forest roads on macroinvertebrate soil fauna of the southern Appalachian mountains, Conservation Biology, 14, 57-63.

Haynes, K.J., Liebhold, A.M., Fearer, T.M., Wang, G., Norman, G.W., Johnson, D.M. (2009) Spatial synchrony propagates through a forest food web via consumer-resource interactions, Ecology, 90, 2974-83.

Haynes, R., Jones, A., Kennedy, V., Harvey, I. & Jewell, T. (2007) District variations in road curvature in England and Wales and their association with road traffic accidents, Environment and Planning, 39, 1222-1237.

Haynes, R., Lake, I. R., Kingham, S., Sabel, C. E., Pearce, J. & Barnett, R. (2008) The influence of road curvature on fatal crashes in New Zealand, Accident Analysis and Prevention, 40, 843-850.

Hels, T. & Buchwald, E. (2001) The effect of road kills on amphibian populations, Biological Conservation, 99, 331-340.

Henriques, L. M. P., Wunderle, J. M. & Willing, M. (2003) Birds of the Tapajos national forest, Brazilian Amazon: A preliminary assessment, Ornitologia Neotropical, 14, 307-338.

Hernandez, M. (1988) Road mortality of the little owl (Athene noctua) in Spain, Journal of Raptor Research, 22, 81-84.

Hilborn, R. &Mangel, M. (1997) The Ecological Detective: Confronting Models with Data.Princeton University Press.

Hindley, G. (1971) A history of roads, Peter Davis, London.

Hobbs, R. J. & Huenneke, L. F. (1992) Disturbance, diversity and invasion: implications for conservation, Conservation Biology, 6, 324-337.

Hodson, N. L. (1962) Some Notes on the Causes of Bird Road Casualties, Bird Study, 9, 168-173.

Honu, Y. A. K. & Gibson, D. J. (2006) Microhabitat factors and the distribution of exotic species across forest edges in temperate deciduous forest of southern Illinois, USA, Journal of the Torrey Botanical Society, 133, 255-266.

Homan, R. N., Windmiller, B. S. & Reed, J. M. (2004) Critical thresholds associated with habitat loss for two vernal pool-breeding amphibians, Ecological Applications, 14, 1547-1553.

Hubbell, S. P., He, F., Condit, R., Borda de Agua, L., Kellner, J. & ter Steege, H. (2008) How many tree species are there in the Amazon and how many will go extinct?, PNAS, 105, 11498-11504.

Hunt, R., Hand, D. W., Hannah, M. A. & Neal, A. M. (1991) Response to CO2 enrichment in 27 herbaceous species, Functional Ecology, 5, 410-421.

Iason, G . R. & Hester, A. J. (1993) The response of heather (Calluna vulgaris) to shade and nutrients- predictions of the carbon- nutrient balance hypothesis, The Journal of Ecology, 81, 75-80.

IBGE (2011) website available from: ftp://geoftp.ibge.gov.br/mapas/banco_dados_georeferenciado_recursos_naturais/ Accessed January 2011.

IFN (2013) http://ifn.florestal.gov.br Accessed June 2013.

IMAZON (2011) http://www.imazon.org.br/programs/landscape-monitoring Accessed May 2013

IPEA (2012) www.ipea.gov.br.

ITTO international Tropical Timber Organisation (2011) website available: http://www.itto.int/ Accessed March 2011.

Jaarsma, C. F., van Langevelde, F. & Botma, H. (2006) Flattened fauna and mitigation: Traffic victems related to road, traffic, vehicle, and species characteristics, Transportation Research Part D 11, 264-276.

Jaeger, J. A. G., Bowman, J., Brennan, J., Fahrig, L., Bert, D., Bouchard, J., Charbonneau, N., Frank, K., Gruber, B., von Toschanowitz, K. T. (2005) Ecological Modelling, 185, 329-348.

Jepson, W. (2006) Private agricultural colonization on a Brazilian frontier, 1970-1980, Journal of Historical Geography, 32, 839-863.

Jiang, Z. 2007. The Road Extension Model in the Land Change Modeler for Ecological Sustainability of IDRISI. Proceedings of the 15th International Syposium on Advances in Geographic Information Systems: 1-8.

Johns, J. S., Barreto, P. & Uhl, C. (1996) Logging damage during planned and unplanned logging operations in the eastern amazon, Forest Ecology and Management, 89, 59-77.

Jones, J. A., Swanson, F. J., Wemple, B. C. & Snyder, K. U. (2000) Effects of roads on hydrology, geomorphology, and disturbance patches in stream networks, Conservation Biology, 14, 76-85.

Jules, E. S., Kauffman, M. J., Ritts, W. D. & Carroll, A. L. (2002) Spread of an invasive pathogen over a variable landscape: a non-native root rot on Port Orford ceder, Ecology, 83, 3167-3181.

Kahm, M., Hasenbrink, G., Lichtenberg-Frate, H., Ludwig, J. & Kschischo, M. (2010) grofit, Fitting Biological Growth Curves with R. Journal of Statistical Software, 33, 1-21. URL http,//www.jstatsoft.org/v33/i07/.

Kalisz, P. J. & Powell, J. E. (2003) Effect of calcareous road dust on land snails (Gastropoda: Pulmonata) and millipedes (Diplopoda) in acid forest soils of the Daniel Boone National Forest of Kentucky, USA, Forest Ecology and Management, 186, 177-183.

Keller, I. & Largiader, C. R. (2003) Recent habitat fragmentation caused by major roads leads to reduction of gene flow and loss of genetic variability in ground b eetles, Proceedings of the Royal Society, 270, 417-423.

Kelly, N. E., Sparks, D. W., DeVault, T. L. & Rhodes, O. E. (2007) Diet of Black and Turkey vultures in a forested landscape, The Wilson Journal of Ornithology, 119, 267-270.

Kerley, L. L., Goodrich, J. M., Miquelle, D. G., Smirnov, E. N., Quigley, H. B. & Hornocker, M. G. (2002) Effects of roads and human disturbance on Amur Tigers, Conservation Biology, 16, 97-108.

Killeen, T.J. 2005.A Perfect Storm in the Amazon Wilderness: Development and Conservation in the Context of the Initiative for the Integration of the Regional Infrastructure of South America (IIRSA). Advances in Applied Biodiversity Series, Number 7, Centre for Applied BiodiversityScience (CABS), Conservation International, USA. Available from : http://www.conservation.org/publications/Documents/AABS.7_Perfect_storm_English.low.res.pdf, Accessed 08/01/2013

King, D. I. &DeGraaf, R. M. (2002) The effect of forest roads on the reproductive success of forest-dwelling passerine birds, Forest Science, 48, 391-396.

Kirby, K. R., Laurance, W. F., Albernaz, A. K., Schroth, G., Fearnside, P. M., Bergen, S., Venticinque, E. M. & da Costa, C. (2006) The future of deforestation in the Brazilian Amazon, Futures, 38, 432-453.

Klepper, O. (1997) Multivariate aspects of model uncertainty analysis: tools for sensitivity analysis and calibration, Ecological Modelling, 101, 1-13.

Kociolek, A. V., Clevenger, A. P., St Clair, C. C. & Proppe, D. S. (2011) Effects of road networks on bird populations, Conservation Biology, 25, 241-249.

Koenker, R. (2010).. quantreg, Quantile Regression. R package, version 4.53. Available from: http,//CRAN.R-project.org/package=quantreg

Koorey, G. (2009) Using an interactive display to demonstrate transportation planning and design issues: getting from A to B, Journal of the Transportation Research Board, 2109, 31-36.

Kriska, G., Bernath, B. & Horvath, G. (2007) Positive Polarotaxis in a mayfly that never leaves the water surface: polarotactic water detection in Palingenia longicauda (Ephemeroptera), Naturwissenschaften, 94, 148-154.

Kriska, G., Bernath, B., Farkas, R. & Horvath, G. (2009) Degrees of Polarization, of Reflected Light Eliciting Polarotaxis in Dragonflies (Odonta), Mayflies (Ephemeroptera) and Tabanid Flies (Tabanidae), Journal of Insect Physiology, 55, 1167-1173.

Kriska, G., Horvath, G. & Andrikovics, S. (1998) Why do Mayflies Lay Their Eggs En Masse on Dry Asphalt Roads? Water-Imitating Polarized Light Reflected From Asphalt Attracts Ephemeroptera, The Journal of Experimental Biology, 201, 2273-2286.

Kristan, W. B., Boarman, W. I. & Crayon, J. J. (2004) Diet composition of common ravens across the urban-wildland interface of the West Mojave desert, Wildlife Society Bulletin, 31, 244-253.

Kuitunen, M., Rossi, E. &Stenross, A. (1999) Do highways influence density of land birds?, Environmental management, 22, 297-302.

Lapola, D. M., Schaldach, R., Alcamo, J., Bondeau, A., Koch, J., Koelking, C., Priess, J. A. (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.

Laporte, N. T., Stabach, J. A., Grosch, R., Lin T. S., & Goetz, S. J. (2007) Expansion of industrial logging in central Africa, Science, 316, 1451.

Laurance, W. F., Albernaz, A. K. M., Schroth, G., Fearnside, P. M., Bergen, S., *et al.* (2002) Predictors of deforestation in the Brazilian Amazon, Journal of Biogeography, 29, 737-748.

Laurance, S. G. (2004) Responses of understory rain forest birds to road edges in central Amazonia, Ecological applications, 14, 1344-1357.

Laurance, S. G., Stouffer, P. C. &Laurance, W. F. (2004) Effects of road clearings on movement patterns of understory rainforest birds in central Amazonia, Conservation biology, 18, 1099-1109.

Laurance, W. F. (2000) Mega-development trend in the Amazon, implications for global change, Environmental Monitoring and Assesment, 61, 113-122.

Laurance, W. F., Goosem, M. & Laurance, G. W. (2009) Impacts of roads and linear clearings on tropical forests, Trends in Ecology and Evolution, 24, 659-669.

Laurance, W.F., Cochrance, M.A., Bergen, S., Fearnside, P.M., Delamonica, P., Barber, C., D'angelo, S & Fernandes, T. (2001) The future of the Brazilian Amazon. Science, 291, 438-439.

Laurance, W. F. & Balmford, A. (2013) A global map for road building, Nature, 495, 308-309.

Laurian, C., Dussault, C., Ouellet, J., Courtois, R., Poulin, M. & Breton, L. (2008) Behaviour of moose relative to a road network, Journal of Wildlife Management, 72, 1550-1557.

LeCorre, M., Ollivier, A., Ribes, S. & Jouventin, P. (2002) Light-induced Mortality of Petrels: A 4-year study from Reunion Island (Indian Ocean), Biological Conservation, 105, 93-102.

Lee, A. C. & Peres (2008) Gap-crossing movements predict species occupancy in Amazonian forest fragments, Oikos, 118, 280-290.

Lees, A. C., Moura, N. G., Andretti, C. B., Davis, B. J. W., Lopes, E.V., Henriques, L. M. P., Aleixo, A. L. P., Barlow, J., Ferreira, J. & Gardner, T. A. (2013). One hundred and thirty-five years of avifaunal surveys around Santarém, central Brazilian Amazonia. RevistaBrasileira de Ornitologia. In press

Legendre , P. & Legendre, L. (1998). Numerical Ecology, 2nd edition. Developments in Environmental Modelling 20. Elsevier.

Lehman, S. M., Rajaonson, A. & Day, S. (2006) Lemur responses to edge effects in the Vohibola III Classifiyed forest, Madagascar, American Journal of Primatology, 68, 293-299.

Lengagne, T. (2008) Traffic noise affects communication behaviour in breeding anuran, Hyla arborea, Biological conservation, 141, 2023-2031.

Lesbarreres, D. & Fahrig, L. (2012) Measures to reduce population fragmentation by roads: what has worked and how do we know?, Trends in Ecology and Evolution, 27, 374-380.

Li, Y. & Briggs, R. (2009) Automatic extraction of roads from igh resolution aerial and satellite images with heavy noise, World Academy of Science, Engineering and Technology, 54, 416-422.

Liu, K. & Sessions, J. (1993) Preliminary planning of road systems using digital terrain models, Journal of Forest Engineering, 4, 27-32.

Loh, J. & Harmon, D. (2005) A global index of biocultural diversity, Ecological Indicators, 5, 231-241.

Ludewigs, T., De Oliveira D'Antona, A., Brondizio, E. S. & Hetrick, S. (2009) Agrarian structure and land-cover change along the lifespan of three colonization areas in the Brazilian Amazon, World Development, 37, 1348-1359.

Lugo, A. E. & Gucinski, H. (2000) Function, effects and management of forest roads, Forest Ecology and Management, 133, 249-262.

MacArthur, R. H., and E. O. Wilson. The the theory of island biogeography: Monographs in Population Biology. Princeton University Press, Princeton, New Jersey (1967).

Mace, R. D., Waller, J.S., Manley, T. L., Lyon, L.J. & Zuuring, H. (1996) Relationships among Grizzly bears, roads and habitat in the Swan mountains, Montana, Journal of Applied ecology, 33, 1395-1404.

Macedo, D. S. & Anderson, A. B. (1993) Early ecological changes associated with logging in an Amazon floodplain, Biotropica, 25, 151-163.

Maeda, E. E., De Almeida, C. M., De Carvalho Ximenes, A., Formaggio, A. R., Shimabukuro, Y. E., Pellikka, P. (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.

Majdi, H. & Persson, H. (1989) effects of road-traffic pollutatnts (lead and cadmium) on tree fine-roots along a motor road, Plant and Soil, 119, 1-5.

Maki, S., Kalliola, R. & Vuorinen, K. (2001) Road construction in the Peruvian Amazon; processes, causes and consequences, Environmental Conservation, 28, 199-214.

Malhi, Y., Baker, T., Wright, J., Phillips, O.L., Almeida, S., *et al.* (2004) The above ground coarse wood productivity of 104 neotropical forest plots, Global Change Biology, 10, 563-591.

Malhi, Y., Phillips, O.L., Lloyd, J., Baker, T., Wright, J., et al, (2002) An international network to monitor the structure, composition and dynamics of Amazonian forests (RAINFOR), Journal of Vegetation Science, 13, 439-450.

Malhi, Y., Wood, D., Baker, T., Wright, J., Phillips, O.L., *et al.* (2006) The regional variation of aboveground live biomass in old-growth Amazonian forests, 12, 1107-1138.

Malhi Y. & Wright J. (2004) Spatial patterns and recent trends in the climate of tropical regions, Philosophical Transactions of the Royal Society London B, 359, 311-329.

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

Malhi, Y., Roberts, T., Betts, R. A., Kelleen, T. J., Li, W. & Nobre, C. A. (2008) Climate change, deforestation and the fate of the Amazon, Science, 319, 169-172.

Manabe, K., Sadr, E. I. & Dooling, R. J. (1998) Control of vocal intensity in budgerigars (Melopsittacus undulatus): Differential reienforcement of vocal intensity and the Lombard effect, Acoustical Scociety of America, 103, 1190-1198.

Mann, M. L., Kaufmann, R. K., Bauer, D., *et al.* (2010) The economics of cropland conversion in Amazonia: The importance of agricultural rent. Ecological Economics, 69, 1503- 1509.

Marini, M. A. & Garcia, F. I. (2005) Bird conservation in Brazil, Conservation Biology, 19, 665-671.

Marsh, D. M., Milam, G. S., Gorham, N. P. & Beckman, N. G. (2005) Forest roads as partial barriers to terrestrial salamander movement, Conservation Biology, 19, 2004-2008

Marsh, D. M. & Beckman, N. G. (2004) Effects of Forest roads on the abundance and activity of terrestrial salamanders, Ecological Applications, 14, 1882-1891.

Matricardi, E. A. T., Skole, D. L., Cochrane, M. A., Qi, J. & Chomentowski, W. (2005) Monitoring selective logging in tropical evergreen forests using Landsat: Multitemporal regional analysis in Matto Grosso, Brazil, Earth Interactions, 9, 1-24.

McGarigal, K., Romme, W. H., Crist, M. & Roworth, E. (2001) Cumulative effects of roads and logging on landscape structure in the San Juan Mountains, Colorado (USA), Landscape Ecology, 16, 327-349.

McGregor, R. L., Bender, D. J. & Fahrig, L. (2008) Do small mammals avoid roads because of the traffic?, Journal of Applied Ecology, 45, 117-123.

McLellan, B. N. & Shackleton, D. M. (1988) Grizzly Bears and resource-extraction industries: Effects of roads on behaviour, habitat use and demography, Journal of Applied Ecology, 25, 451-460.

MEA (Millennium Ecosystem Assessment), Current state and trends (2005) Available from http://www.maweb.org/en/Condition.aspx.

Mech, D. L., Fritts, S. H., Radde, G. L. & William, P. J. (1988) Wolf Distribution and Road Density in Minnesota, Wildlife Society Bullitin, 16, 85-87.

Melillo, J. M., Houghton, R. A., Kicklighter, D. W. & McGuire, A. D. (1996) Tropical deforestation and the global carbon budget, Annual Review Energy and Environment, 21, 293-310.

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

Mena, J. B. (2003) State of the art on automatic road extraction for GIS update,a novel classification, Pattern Recognition Letters, 24, 3037-3058.

Merry, F., Soares-Filho, B., Nepstad, D., Amacher, G., Rodrigues, H. (2009) Balancing conservation and economic sustainability: the future of the Amazon timber industry, Environmental Management, 44, 395-407.

Merry, F. D. & Amacher, G. S. (2005) Forest taxes, timber concessions and policy choices in the Amazon, Journal of Sustainable Forestry, 20, 15-44

Mertens, B., Poccard-Chapuis, R., Piketty, M. G., Lacques, A. E. & Venturieri, A. (2002) Crossing spatial analyses and livestock economics to understand deforestation processes in the Brazilian Amazon: the case of Sao Felix do Xingu in south Para, Agricultural Economics, 27, 269-294.

Messina, J. P. & Walsh, S. J. (2001) 2.5D morphogenesis, modelling landuse and landcover dynamics in the Ecuadorian Amazon, Plant Ecology, 156, 75-88.

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

Mineau, P. & Brownlee, L. J. (2005) Road salts and birds: an assessment of the risk with particular emphisis on winter finch mortality, Wildlife Society Bulletin, 33, 835-841.

Minunno, F, Oijen, M. V., Cameron, D. R., Cerasoil, S., Pereira, J. S. & Tome, M. (2013) Using a Bayesian framework and global sensitivity analysis to idensify strengths and weaknesses of two process based models differing in represesntation of autotrophic respiration, Environmental Modeling and Software, 42, 99-115.

Montagnini, F. & Jordan, C. F. (2005) Tropical Forest Ecology, Springer, Berlin.

Montanari, A. (2007) What do we mean by uncertainty? The need for consistant wording about uncertainty assessment in hydrology, Hydrological Processes, 21, 841-845.

Moore, R. P., Robinson, W. D., Lovette, I. J. & Robinson, T. R. (2008) Experimental evidence for extreme dispersal limitation in tropical forest birds, Ecology letters, 11, 960-968.

Moran , E. F. (1993) Deforestation and land use in the Brazilian Amazon, Human Ecology, 21, 1-21.

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

Morton, D. C., R. DeFries, Y. E. Shimabukuro, L. O. Anderson, E. Arai, F. d. B. 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 U.S.A. 103, 14637-14641.

Movaghati, S., Moghaddamjoo, A. & Tavakoli, A. (2010) Road extraction from satellite images using particle filtering and extended Kalman filtering, IEEE Transactions on Geoscience and Remote Sensing, 48, 2807-2817.

Muller, R., Muller, D., Schierhorn, F. & Gerold, G. (2010) Spatiotemporal modeling of the expansion of mechanised agriculture in the bolivian lowlad forests, Applied Geography, 31, 631-640.

Mumme, R. L., Schoech, S. J., Woolfenden, G. E. & Fitzpatrick, J. W. (2000) Life and death in the fast lane: demographic consequences of road mortality in the Florida Scrub-Jay, 14, 501-512.

Munguira, M. L. & Thomas, J. A. (1992) Use of road verges by butterfly and burnet populations and the effect of roads on adult dispersal and mortality, Journal of applied Ecology, 29, 316-329.

Munnell, A. H. (1992) Infrastructure investment and economic growth, The journal of Economic Prespectives, 6, 189-198.

Munro, K. G., Bowman, J. & Fahrig, L. (2012) Effect of paved road density on abundance of white-tailed deer, Wildlife Research, 39, 478-487.

Murcia, C. (1995) Edge effects in fragmented forests: implications for conservation, Trends in Ecology and Evolution, 10, 58-62.

Murray, J. D. (2002) Mathematical Biology: I. An Introduction. Verlag: Springer.

Nepstad, D., Merry, F., Rodrigues, H. O., Schwartzman, S. (2007) The costs and benefits of reducing carbon emissions from deforestation and forest degradation in the Brazilian Amazon, REDD, United Nations Framework Convention on Climate Change (UNFCCC), Conference of the Parties (COP), Thirteenth Session. Available online at: http://www.whrc.org/policy/pdf/cop13/WHRC_Amazon_REDD.pdf Accessed: 2012 Apr 2.

Nepstad, D., Soares-Filho, B. S., Merry, F., Lima, A., Moutinho, P., *et al.* (2009) The end of deforestation in the Brazilian Amazon, Science, 326, 1350-1351.

Nepstad, D. C., Verissimo, A., Alencar, A., Nobre, C., Lima, E., Lefebvre, P., Schlesinger, P., Potter, C., Moutinho, P., Mendoza, E.,Cochrane, M. & Brooks, V. (1999) Large scale impoverishment of Amazonian forests by logging and fire, Nature, 398, 505-508.

Nepstad, D., Carvalho, G., Barros, A. C., Alencar, A., Capobianco, J. P., Bishop, J., Moutinho, P., Lefebvre, P., Solva, U. L. & Prins, E. (2001) Road paving, fore regime feedbacks and te future of the Amazon, Forest ecology and anagement, 154, 395-407.

O'Neill, R. V. (1976) Ecosystem persistence and heterotrophic regulation, Ecology, 57, 1244-1253.

Ortega, Y.K. & Capen, D. E. (1999) Effects of forest roads on habitat quality for ovenbirds in a forested landscape, The Auk, 116, 937-946.

Ortowski, G. & Nowak, L. (2006) Factors influencing mammal roadkills in the agricultural landscape of south-western Poland, Polish Journal of Ecology, 54, 283-294.

Ortowski, G. (2008) Roadside Hedgerows and Trees as Factors Increasing Road mortality of Birds: Implications for management of Roadside Vegitation in rural landscapes, Lnadscape and urban Planning, 86, 153-161.

Parendes, L. A. & Jones, J. A. (2000) Role of light availability and dispersal in exotic plant invasion along roads and streams in the H. J. Andrews Experimental forest, Oregon, Conservation Biology, 14, 64-75.

Parris, K. & Schneider, A. (2009) Impacts of traffic noise and traffic volume on birds of roadside habitats, Ecology & Society, 14, 29.

Pelt Pará (2012) Available from: http://www.setran.pa.gov.br/PELT/tranporte/arquivos/evol_futura_transp.pdf

Pereira, R., Zweede, J., Asner, G. P. & Keller, M. (2001) Forest canopy damage and recovery in reduced-impact and conventional logging in eastern Para, Brazil, forest ecology and management, 5778, 1-13.

Peres, C. A., Barlow, J. &Laurance, W. F. (2006) Detecting antropogenic disturbance in tropical forests, TREE, 21, 227-229.

Perz, S. G., Cabrera, L., Carvalho, L. A., Castillo, J., Chacacanta, R., Cossio, R. E., Solano, Y. F., Hoelle, J., Perales, L. M., Puerta, I., Cespedes, D. R., Camacho, I. R. & Silva, A. C. (2012) Regional integration and local change: road paving, community connectivity and social-ecological resilience in a tri-national frontier, southwestern Amazonia, Regional Environmental Change, 12, 35-53.

Perz, S. G., Caldas, M. M., Arima, E. & Walker, R. J. (2007) Unofficial road building in the Amazon: socioeconomic and biophysical explanations, Development and Change, 38, 529-551.

Perz, S., Brilhante, S., Brown, F., Caldas, M., Ikeda, S., Mendoza, E., Overdevest, C., Reis, V., Reyes, J.F., Rojas, D., Schmink, M., Souza, C. & Walker, R. (2008) Road building, land use and climate change, prospects for environmental governance in the Amazon, Philosophical Transactions of the Royal Society B, 363, 1889-1895.

Phillips, O. L., Baker, L., Arroyo, L., Higuchi, N., Killeen, T. J., *et al.* (2004) Pattern and process in Amazon tree turnover 1976-2001, Philosophical Transactions of the Royal Society London B, 359, 381-407.

Pickles, W. (1942) Animal Mortality on Three Miles of Yorkshire Roads, Journal of Animal Ecology, 11, 37-43.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and the R Development Core Team (2012). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-106.

Pontius, R. G. Jr. (2002) Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions.Photogrammetric Engineering & Remote Sensing 68(10) p.1041-1049.

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

Potvin, M.J, Drummer, T.D., Vucetich, J.A., Beyer, D. E., Peterson, R.O., Hammill, J. H. (2005) Monitoring and habitat analysis for wolves in upper Michigan, Journal of Wildlife Management, 69, 1660-1669.

Prates-Clark, C. D. C., Saatchi, S. S. & Agosti, D. (2008) Predicting geographical distribution models of high value timber trees in the Amazon basin using remotely sensed data, Ecological Modelling, 211: 309-323.

Pytte, C. L., Rusch, K. M. & Ficken, M. S. (2003) Regulation of vocal amplitude by the Blue-throated hummingbirds, Lampornis clemenciae, Animal Behaviour, 66, 703-710.

Quesada, C. A., Lloyd, J., Schwarz, M., Baker, T., Phillips, O. L., *et al.* (2009) Regional and large scale patterns in Amazon forest structure and function are mediated by variations in soil physical and chemical properties, Biogeosciences Discussions, 6, 3993-4057.

Quinn, J. L., Whittingham, M. J., Butler, S. J. & Cresswell, W. (2006) Noise, predation reisk compensation and vigilance in the chaffinch Fringilla coelebs, Journal of Avian Biology, 37, 601-608.

R Development Core Team (2009) R, A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Radford, J. Q., Bennett, A. F., Cheers, G. J. (2005) Landscape level thresholds of habitat cover for woodland dependant birds, Biological Conservation, 124, 317-337.

Ralston, B. A. and Barber, G. M. 1982. A Theoretical Model of Road Development Dynamics. Annals of the association of american geographers, 72, 201-210.

Ramos, A., R. (1998) Indigenism: ethnic politics in Brazil, The University of Wisconsin Press, Wisconsin.

Refsgaard, J. C., Sluijs, J. P. V., Hojberg, A. L. & Vanrolleghem, P. A. (2007) Uncertanty in the environmental modelling process a framework and guidance, Environmental Modelling and Software, 22, 1543-1556.

Regan, H. M., Colyvan, M. & Burgman, M. A. (2002) A taxonomy and treatment of uncertainty for ecology and conservation biology, Ecological Applications, 12, 618-628.

Reijnen, R. & Foppen, R. (1991) Effect of road traffic on the breeding site-tenacity of male Willow Warbles (Phylloscopustrochilus), Journal of ornithology, 132, 291-295.

Rene, L. 1964. Networks and the location of economic activities. Reginal Science Association, 183-196.

Richard, Y. & Armstrong, D. P. (2010) Cost distance modelling of landscape connectivity and gap-crossing ability using radio-tracking data, Journal of Applied Ecology, 47, 603-610.

Richardson, J. H., Shore, R. F., Treweek, J. R. & Larkin, S. B. C. (1997) Are major roads a barrier to small mammals?, Journal of Zoology, 243, 840-846.

Rico, A., Kindlmann, P. & Sedlacek, F. (2007) Barrier effects of roads on movments of small mammals, Folia Zoologica, 56, 1-12.

Rodrigues, A. S. L., Ewers, R. M., Parry, L., Souza, C. Jr., Verissimo, A. & Balmford, A. (2009) Boom-and-bust development patterns across the Amazon deforestation frontier, Science, 324, 1435-1437.

Rodrigues-Filho, S., Bursztyn, M., Lindoso, D., Debortoli, N., Nesheim, I. & Verberg, R. (2012) Road development and deforestation in Amazonia, Brazil, In: McNeill, D., Nesheim, I. & Brouwer, F. (2012) Land use policies for sustainable development: Exploring integrated assessment approaches, Edward Elgar, Northampton, MA, USA.

Roger, E., Laffan, S. W. & Ramp, D. (2010) Road impacts a tipping point for wildlife populations in threatened landscapes, Population Ecology,

Roh, T., Seo, D. & Lee, J. (2003) An accuracy analysis for horizontal alighnment of road by the kinematic GPS/GLONASS combination, KSCE Journal of Civil Engineering, 7, 73-79.

Rosa I.M.D., Ahmed S.E. & Ewers R.M. (2014 In Press). The transparency, reliability and utility of land-use and land-cover change models (Appendix D).

Row, J. R., Blouin-Demers, G. & Weatherhead, P. J. (2007) Demograpgic effects of road mortality in black ratsnakes (Elaphe obsoleta), Biological Conservation, 137, 117-124.

Rydell, J. (1992) Exploitation of insects around streetlamps by bats in Sweden, Functional Ecology, 6, 744-750.

Rytwinski, T. & Fahrig, L. (2007) Effect of road density on abundance of white-footed mice, Landscape Ecology, 22, 1501-1512.

Rytwinski, T. & Fahrig, L. (2011) Reproductive rate and body size predict road impacts on mammal abundance, Ecological Applications, 21, 589-600.

Saatchi, S. S., Houghton, R. A., Dos Santos Alvala, R. C., Soares, J. V., Yu, Y. (2007) Distribution of aboveground live biomass in the Amazon Basin, Global Change Biology, 13, 816-837.

Sahin, V. & Hall, M. J. (1996) The effects of afforestation and deforestation on water yields, Journal of Hydrology, 178, 293-309.

Santos, C. D., Miranda, A. C., Granaderia, J. P., Lourenco, P. M., Saraiva, S. & Palmeirim, J. M. (2010) Effects of Artificial Illumination on the Nocturnal Foraging of Wader, Acta Oecologica, 36, 166-172.

Sanzo, D. & Hecnar, S. J. (2006) Effects of de-icing salt (NaCl) on Larval wood frogs (Rana sylvatica), Environmental Pollution, 140, 247-256.

Saunders, S. C., Mislivets, M. R., Chen, J. & Cleland, D. T. (2002) Effects of roads on landscape structure within nested ecological units of the Northern Great Lakes Region, USA, Biological Conservation, 103, 209-225.

Sawyer, H., Lindzey, F. & McWhirter, D. (2005) Mule deer and pronghorn migration in western Wyoming, Wildlife Society Bulletin, 33, 1266-1273.

Schaub, A., Ostwald, J. & Siemers, B. M. (2008) Foraging Bats Avoid Noise, The Journal of Experimental Biology, 211, 3174-3180.

Scott, T. G. (1938) Wildlife Mortality on Iowa Highways, American Midland Naturalist, 20, 527-539.

Seiler, A., Helldin, J. O. & Seiler, c. (2004) Road mortality in Swedish mammals: results of a drivers' questionare, Wildlife Biology, 10, 225-233.

Semlitsch, R. D., Ryan, T. J., Hamed, K., Chatfield, M., Drehman, B., Pekarek, N., Spath, M. & Watland, A. (2007) Salamander abundance along road edges and within abandpned logging roads in Appalachian forests, Conservation biology, 21, 159-167.

Sevstuk, A. (2010) Path and place: a study of urban geometry and retail activity in Cambridge and Somerville, MA, PhD Dissertaion, MIT, Cambridge.

SFB (2013) Servico Florestal Brazileiro (Service Forest Brazil) http://www.florestal.gov.br/concessoes-florestais/florestas-sob-concessao/duas-florestas-nacionais-abrigam-concessao-florestal Accessed 05/04/2013.

Shepard, D. B., Kahns, A. R., Dreslik, M. J. & Phillips, C. A. (2008) Roads as barriers to animal movement in fragmented landscapes, Animal Conservation, 11, 288-296.

Sherratt, J. A. & Smith, M. J. (2008) Periodic travelling waves in cyclic populations: field studies and reaction-diffusion models, Journal of the Royal Society Interface, 5, 483-505.

Shyama Prasad Rao, R. & Saptha Girish, M. K. (2007) Road kills: assessing insect casualties using flagship taxa, Current Science, 92, 830-837.

Sierra, R. (2001) The role of domestic timber markets in tropical deforestation and forest degredation in Ecuador: Implications for conservation planning and policy, Ecological Economics, 36, 327-340.

Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. (2005) ROCR: visualizing classifier performance in R. Bioinformatics 21(20):3940-3941.

Slabbekoon, H. & Ripmeester, E. A. P. (2008) Birdsong and anthropgenic noise: implications and applications for conservation, Molecular Ecology, 17, 72-83.

Smith, G. D. (1986) Numerical Solution of Partial Differential Equations: Finite Difference Methods. Oxford University Press, Oxford UK.

Soares-Filho, B. S., Cerqueira, G. C. & Pennachin, C. L. (2002) Dinamica a stochastic cellular automata model designed to simmulate landscape dynamics in an amazon colonisation frontier, Ecological Modelling, 154, 217-235.

Soares-Filho, B., Alencar, A., Nepstad, D., Cerqueira, G., Diaz, M.D.C.V., Rivero, S., Solorzano, L. & Voll, E. (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., Nepstead, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., Voll, E., McDonald, A., Lefebvre, P. & Schelsinger, P. (2006) Modelling conservation in the Amazon basin, Nature, 440, 520-523.

Southworth, J., Marsik, M., Qiu, Y., Perz, S., Cumming, G., Stevens, F., Rocha, K., Duchelle, A. & Barnes, G. (2011) Roads as drivers of change: trajectories across the Tri-national frontier in MAP, the Southwestern Amazon, Remote Sensing, 3, 1047-1066.

Spellberg, I. F. (1998) Ecological effects of roads and traffic: A literature review, Global Ecology and Biogeography Letters, 7, 317-333.

Spellerberg, I. F. (2002) Ecological Effects of Roads, Science Publishers, Inc., Enfeld, USA.

Spooner, G. P. & Smallbone, L. (2009) Effects of road age on the structure of roadside vegetation in south-eastern Australia, Agriculture, Ecosystems and Environment, 129, 57-64.

Spracklen, D. V., Arnold, S. R. & Taylor, C. M. (2012) Observations of increased tropical rainfall preceded by air passage over forests, Nature, 489, 282-286.

Stapp, P. & Lindquist, M. D. (2007) Roadside foraging by kangaroo rats in a grazed short-grass prairie landscape, Western North American Naturalist, 67, 368-377.

Steffan-Dewenter, I. (2002) Importance of habitat area and landscape context for species richness of bees and wasps in fragmented orchard meadows, Conservation Biology, 17, 1036-1044.

Steffan-Dewenter, I., Munzenberg, U., Burger, C., Thies, C. & Tscharntke, T. (2002) Scale dependant effects of landscape context on three pollinator guilds, Ecology, 83, 1421-1432.

Stephenson , N. L. & Van Mantgem, P. J. (2005) Forest turnover rates follow global and regional patterns of productivity, Ecology Letters, 8, 524-531.

Stone, S. W. (1998) Using a geographic information system for applied policy analysis: the case of logging in the Eastern Amazon, Ecological Economics, 27, 43-61.

Stoner, D. (1935) Highway Mortality among Mammals, Science, 81, 401-402.

Storch, H. V., Costa-Cabral, M., Hagner, C., Feser, F., Pacyna, J., Pacyna, E. & Kolb, S. (2003) Four decades of gasoline lead emissions and control policies in Europe: a retrosprctive assessment. The Science of the Total Environment, 311, 151-176.

Straub, S. (2008) Infrastructure and growth in developing countries: recent advances and research challenges, World bank policy research working paper, number 4460, available from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1080475 (accessed 11/12/2012).

Suarez, A. V., Bolger, D. T. & Case, T. J. (1998) Effects of fragmentation and invasion on native ant communities in coastal southern California, Ecology, 79, 2041-2056.

Swanepoel, J. W., Kruger, G. H. J. & van Heerden, P. D. R. (2007) Effects of sulphur dioxide on photosynthesis in the succulent Augea capensis Thunb, Journal of Arid Environments, 70, 208-221.

Taaffe, E. J., Morrill, R. L. & Gould, P. R. (1963) Transport expansion in underdeveloped countries; a comparative analysis, American Geographical Society, 53, 503-529.

Taiz, L. & Zeiger, E. (2006) Plant physiology, Fourth edition, Sinauer Associates Inc. Publishers, Sunderland, USA.

Taylor, B. D. & Goldingay, R. L. (2012) Restoring connectivity in landscapes fragmented by major roads: A case study using wooden poles as ‘stepping stones’ for gliding mammals, Restoration Ecology, 20, 671-678.

ter Steege H, Pitman NCA, Phillips OL, Chave J, Sabatier D, *et al.* (2003) A spatial model of tree α-diversity and tree density for the Amazon, Biodiversity and Conservation, 12: 2255-2277.

ter Steege H, Pitman NCA, Phillips OL, Chave J, Sabatier D, *et al.* (2006) Continental-scale patterns of canopy tree composition and function across Amazonia, Nature, 443, 444-447.

Thiel, R. P. (1985) Relationship between road densities and wolf habitat suitability in Wisconsin, American Midland Naturalist, 113, 404-407.

Theiler, J. (1990) Estimating Fractal Dimension, Journal of the Optical Society of America, 7, 1055-1073.

Tremblay, M. A. & St. Clair, C. C. (2009) Factors affecting the permeability or transportation and riparian corridors to the movements of songbirds in an urban landscape, Journal of Applied Ecology, 46, 1314-1322.

Trombulak, S. C. & Frissell, C. A. (2000) Review of ecological effects of roads on terrestrial and aquatic communities, Conservation Biology, 14, 18-30.

Tuxbury, S. M. & Salmon, M. (2005) Competitive Interactions Between Artificial Lighting and Natural Cues During Seafinding by Hatchling Marine Turtles, Biological Conservation, 121, 311-316.

Uhl C, Barreto P., Verissimo A., Vidal E., Amaral P., *et al.* (1997) Natural resource management in the Brazilian Amazon, BioScience, 47, 160-168.

Uhl, C. & Guimaraes Vieira, I. C. (1989) Ecological impacts of selective logging in the Brazilan Amazon: a case study from the Paragominas region of the state of Para, Biotropica, 21, 98-106.

Uhl, C., Verissimo, A., Mattos, M. M., Brandino, Z. & Vieira, I. C. G. (1991) Social, economic and ecological consequencesof selective logging in an Amazoninan frontiers: the case of Tailanda, Forest Ecology and Management, 46, 243-273.

UNEP GLOBIO 2001 report available from: http://www.globio.info/downloads/218/globioreportlowres.pdf accessed 02/03/12.

Uriarte, M., M. Pinedo-Vasquez, R. S. DeFries, K. Fernandes, V. Gutierrez-Velez, W. E. Baethgen, and C. Padoch. (2012) Depopulation of rural landscapes exacerbates fire activity in the western Amazon. Proceedings of the National Academy of Sciences 109, 21546-21550.

Van der Zande, A. N., Keurs, W. J. & van der Weilden, W. J. (1980) The impact of roads on the densities of four bird species in an open feild habitat- evidence of a long distance effect. Biological Conservation, 18, 299-321.

Verissimo, A. & Cochrane, M. A. (2003) Brazil’s bold initiative in the Amazon, ITTO Tropical Forest Update, 13, 4-6.

Verissimo, A., Barreto, P., Tarifa, R. & Uhl, C. (1995) Extraction of a high-value natural resource in Amazonia: the case of mahogany, Forest Ecology and Managemt, 72, 39-60.

Verissimo, A., Chochrane, M. A., Souza, C. & Salomao, R. (2002) Priority areas for establishing national forests in the Brazillian Amazon, Conservation Ecology, 6, 4-13

Vijayakumar, S. P., Vasudevan, K. & Ishwar, N. M. (2001) Herpetofaunal mortality on roads in the Anamalai hills, Southern Western Ghats, Hamadryad, 26, 265-272.

Vos, C. C. & Chardon, J.P. (1998) Effects of habitat fragmentation and road density on the distribution pattern of the moor frog rana arvalis, Journal of applied ecology, 35, 44-56.

Walker, R., Arima, E., Messina, J., Soares-Filho, B., Perz, S., Vergara, D., Sales, M., Pereira, R. & Castro, W. (2013) Modeling spatial decisiosns with graph theory; logging roads and forest fragmentation in the Brazilian Amazon, Ecological Applications, 23, 239-254.

Walker, R., Drzyzga, S. A., Li, Y., Qi, J., Caldas, M., Arima, E. & Vergara, D. (2004) A behavioral model of landscape change in the Amazon, basin, the colonist case, Ecological Applications, 14, 299-312.

Walker, W. E., Harremoes, P., Rotmans, J., Sluijs, J. P. V., Van Asselt, M. B. A., Janssen, P. & Von Krauss, M. P. K. (2003) Defining uncertainty, Integrated Assesment, 4, 5-17.

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

Walter, C. & Maguire, J. (1996) Lessons for stock assessment from the northern cod collapse, Reviews in fish biology and fisheries, 6, 125-137.

Wang, F., Mladenoff, D. J., Forrester, J. A., Keough, C. & Patron, W. J. (2013) Global sensitivity analysis of a modifiyed CENTURY model for simulating impacts of harvesting woody biomass for bioenergy, Ecological Modelling, 259, 16-23.

Warren, P. S., Katti, M., Ermann, M. & Brazel, A. (2006) Urban bioacoustics: it's not just noise, Animal Behaviour, 71, 491-502.

Wassenaar, T., Gerber, P., Verburg, P. H., Rosales, M., Ibrahim, M. & Steinfeld, H. (2007) Projecting land use changes in the neo-tropics, the geography of pasture expansion into forest, Global Environmental Change, 17, 86-104.

Watkins, R. Z., Chen, J., Pickens, J. & Brosofske, K. D. (2003) Effects of forest roads on understory plants in a managed hardwood landscape, Conservation Biology, 17, 411-419.

Watts, R. D., Compton, R. W., McCammaon, J. H., Rich, C. L., Wright, S. M., Owens, T. &Ouren, D. S. (2007) Roadless space of the conterminous United States, Science, 316, 736-738.

Wearn, O. R., Reuman, D. C. & Ewers, R. M. (2012) Extinction debt and windows of conservation opportunity in the Brazilian Amazon, Science, 337, 228-232.

Weins, J. A., Moss, M. R., Turner, M. G. & Mladenoff, D. J. (2007) Foundation Papers in Landscape Ecology, Columbia University Press, New York.

Whitmore, T. C. (1998) An introduction to tropical rainforests, Second Edition, Oxford University Press, New York.

Whittington, J., StClair, C.C. &Mercer,G. (2005) Spatial responses of wolves to roads and trails in mountain valleys, Ecological Applications, 15, 543-553.

Wilkie, D., Shaw, E., Rotberg, F., Morelli, G. & Auzel, P. (2000) Roads, development and conservation in the Congo basin, Conservation Biology, 14, 1614-1622.

Witmer, G. W. & deCalesta, D. S. (1985) Effect of Forest Roads on Habitat Use by Roosevelt Elk, Northwest Science, 59, 122-125.

World Bank. (2011) Road density data, http://data.worldbank.org/indicator/IS.ROD.DNST.K2 Accessed 2011 and Dec 2013.

Wright, S. J. & Muller-Landau, H. C. (2006) The future of tropical forest species, Biotropica, 38, 287-301.

WWF (2010) Amazon alive; a decade of discovery 1999-2009, Available at: http://www.worldwildlife.org/what/wherewework/amazon/WWFBinaryitem18397.pdf accessed 06/06/2011.

Yamada, Y., Sasaki, H. & Harauchi, Y. (2010) Effects of narrow roads on the movement of carabid beetles (Coleoptera, Carabidae) in Nopporo Forest Park, Hokkaido, Journal of Insect Conservation, 14, 151-157.

Young, K. R. (1994) Roads and environmental degradation of tropical montane forests, Conservation Biology, 8, 972-976.

Zernitz, E. R. (1932) Drainage patterns and their significance, The Journal of Geology, 40, 498-521.

Acknowledgements quotation;

Excerpt from a letter by Alfred Russel Wallace to friends at the Mechanic Institute at Neath, 1849. Taken from 'Natural selection and beyond; the intellectual legacy of Alfred Russel Wallace'. Edited by Smith, C. H. & Beccaloni, G. (2008) Oxford University Press.

Appendices

Appendix A: **Model code for Chapter 6**

Model Code for Chapter 5

```
// Written in C for Visual Studio
// With thanks to Matthew Smith
// Copyright (C) Microsoft. All rights reserved.

using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;

/// Non standard libraries - in this case Scientific DataSet libraries that
/// allow us to read in and output datasets in various formats (mostly .csv files
/// or .nc files)
using Microsoft.Research.Science.Data;
using Microsoft.Research.Science.Data.CSV;
using Microsoft.Research.Science.Data.Imperative;
using FilzbachInterop;

namespace FittingRoadDynamics
{
  class ModelCode
  {
   // Declare the data path to use for input and output
    string DataPath = "TYPE THE FULL FILE PATH HERE";
    string InputDataFile = "raw_data_Grid_100.csv"; // The input datafile name
    string OutputDataFile = "OutputDataGrid1.csv"; // The output datafile name
    public string RunID = "";// to append to files to indicate which run performed
    string AnalysisName = ""; // also to append to files to identify a paricular experiment

    // Read in the datafile and the data
    public double[] RoadData2004; // Road density in 2004
    public double[] RoadData2007; // Road density in 2007
    public double[] DummyData2007; // for faking the data.
    public double[] Location; // for uniquely identifying each site
    public double[] BackupData;
    public double[] ImedData1;
    public double[] ImedData2;
    public double[] xCoords; // x coordinates
    public double[] yCoords; // y coordinates
    public double[] Coded; // indicates whether the cell density should be simulated
    public const int NFolds = 10; // Number of folds to run
    public static int FoldNumber = 0; // Intialise the fold number
    public static bool Training = true; // A flag to indicate if we are in teh trainign or validation step
    public int[] FoldIndex; // stores the folds associated with each data point
    public static double CurrentLikelihood = 0; // Tracks the current likelihood
    //int NumNeighbours = 13; // The numebr of nearest neighbours to consider in the
neighbourhood model
    public int[] NumNeighbours; // Indicates the number of neighbours for a given site
    public int[,] NeigboursList; // List of nearest neighbours for each site
```

```
        public bool[] IsValid; // Keeps track of whether a datapoint is a valid datapoint - used to exclude
exceptions
        public double[,] EuclideanDistList; // Distance of nearest negibours for each site
        public string Model = "Dispersal1"; // To change the model change this to either "Exponential",
"Logistic", "Dispersal", "Dispersal2" or "Wave"
        public int DisperalSubmodel = 1; // If you pick "Dispersal" or "Dispersal2" then this picks the
submodel
        public Random OurRandomNumberGenerator;
        public static int CountZero = 0; // This is used in computation to keep track of the number of
zero elements
        public static bool FakeDataSet = false; //Indicates whether to fake the dataset being predicted
        public static string ModelToFake = "Logistic"; //Indicates the model to fake
        public const double CodedThresh = 1.1;

        public int[] Quadrant; // If using the travelling wave model then this is the quadrant that the
site has been assigned to
        public double[] CentroidX; // This is the x centroid of the quadrant
        public double[] CentroidY; // This is the y centroid of the quadrant


        public static void MainProcedure(string Model, int DispersalSubmodel, string NameOfAnalysis)
        {
            Console.WriteLine("Program has started"); // Flag to the user that the program initialised
successfully.

            // Points to the function used to assess likelihood
            // It is unlikely that you will need to change these two lines
            // but to use Filzbach using C# we need to create a Program "object"
            ModelCode FittingProgram = new ModelCode();
            Filzbach.pfn_likelihood = FittingProgram.CalculateLikelihoodFilzbach;
            FittingProgram.OurRandomNumberGenerator = new Random(1);
            FittingProgram.Model = Model;
            FittingProgram.DisperalSubmodel = DispersalSubmodel;
            FittingProgram.AnalysisName = NameOfAnalysis;
            FittingProgram.RunID = NameOfAnalysis + FittingProgram.Model +
FittingProgram.DisperalSubmodel.ToString();
            FittingProgram.OutputDataFile = FittingProgram.RunID + "OutputData.csv";

            // This function reads in the data from the datafile
            FittingProgram.ReadInData();

            // This function reads in the data from the datafile
            FittingProgram.AssignFolds();

            // This sets up the output file to take the estimate from the different model folds
            FittingProgram.SetupOutput();

            //This calculates the nearest neaihbours and their distances for the dispersal model, or the
quadrants and centroids for the wave model
            FittingProgram.AssignNeighbours();
```

```
        // Let's fake some data and see if we can back infer what we had
        if (FakeDataSet) FittingProgram.FakeData();

        for (int Fold = 0; Fold < NFolds; Fold++)
        {
          // Initialise Filzbach
          // This function puts a Filzbach "object" in memory
          Filzbach.initialize_filzbach(); // needs to be done to reset Filzbach and remove previous
parameters and values from memory

          // Sets up the parameters
          FittingProgram.SetupParametersToEstimate();

          // Run Filzbach
          Filzbach.set_chains(1); // currently needed to avoid bugs - this defines how many Markov
chains we want to run in parallel
          // This has the form "burnin length", "Chain Length", "Burnin length for Maximum
Likelihood estimate", and "Chain Length"
          FoldNumber = Fold;
          Filzbach.runmcmc(10000, 100000, 1000, 1000); // the main command to run MCMC model
fitting

          // This then puts a variety of outputs to a datafile
          FittingProgram.OutputDataToFileFolds();
        }
      }

      private void FakeData()
      {
        int NumData = RoadData2004.Length; // Work out how much data we need
        DummyData2007 = new double[NumData]; // Stores the list of neigbours
        int iterations = 3000000;
        double r = 0.08;
        double K = 0.5;
        double Kd = 0.3;
        double d = 0.01;
        double dt = 3.0/(double)iterations;
        double var = 0.7;
        double Thresh = 0.0;

        for (int ii = 0; ii < NumData; ii++)
        {
          DummyData2007[ii] = RoadData2004[ii];
        }

        for (int ii = 0; ii < NumData; ii++)
        {
          if (ModelToFake == "Exponential")
          {
            DummyData2007[ii] = ExpFunction(RoadData2004[ii], r, 3.0);
          }
```

```
            else if (ModelToFake == "Logistic")
            {
               DummyData2007[ii] = LogisticFunction(RoadData2004[ii], K, r, 3.0);
            }
            else if (ModelToFake == "Dispersal" || ModelToFake == "Dispersal1")
            {
               DummyData2007[ii] = DispersalFunction(RoadData2004[ii], K, r, d, Thresh, ii, 3.0);
            }
            else if (ModelToFake == "Dispersal2")
            {
               DummyData2007[ii] = DispersalFunction2(RoadData2004[ii], K, r, d, Thresh, ii, 3.0);
            }
            else if (ModelToFake == "Dispersal3")
            {
               DummyData2007[ii] = DispersalFunction3(RoadData2004[ii], K, r, d, Thresh, Kd, ii, 3.0);
            }

            double imed1 = Filzbach.normal_draw(Math.Log10(DummyData2007[ii] -
RoadData2004[ii]), var);
            DummyData2007[ii] = RoadData2004[ii] + Math.Pow(10, imed1);
         }

         for (int ii = 0; ii < NumData; ii++)
         {
            RoadData2007[ii] = DummyData2007[ii];
         }

      }

      /// <summary>
      /// Code to identify the neigbours, to find nearest neighbours based on Euclidean distance
      /// and to idenify the sites according to quadrants
      /// </summary>
      public void AssignNeighbours()
      {
         int NumData = RoadData2004.Length; // Work out how much data we need
         NumNeighbours = new int[NumData]; // Stores the distance of those neighbours
         CentroidX = new double[4]; // Stores the x centroid of each quadrant
         CentroidY = new double[4]; // Stores the y centroid of each quadrant

         DataSet RoadDataFile = DataSet.Open(DataPath + InputDataFile);
         double[,] TempNeighbourList = RoadDataFile.GetData<double[,]>("neighbours"); // get the
neighbour id matrix
         double[] TempNumNeighbours = RoadDataFile.GetData<double[]>("numneighbours"); // get
the count of neighbours
         double[] TempQuadrant = RoadDataFile.GetData<double[]>("quadrant"); // get the quardant
identifier
         RoadDataFile.Dispose();
         int MaxNeighbours = TempNeighbourList.GetLength(1);
         NeigboursList = new int[NumData, MaxNeighbours]; // Stores the list of neigbours
         EuclideanDistList = new double[NumData, MaxNeighbours]; // Stores the list of neigbours
```

```
        for (int iD = 0; iD < NumData; iD++) // for each site
        {
           NumNeighbours[iD] = (int)TempNumNeighbours[iD];
           for (int iD2 = 0; iD2 < MaxNeighbours; iD2++) // work out distance to site
           {
              NeigboursList[iD, iD2] = (int)TempNeighbourList[iD, iD2];
              EuclideanDistList[iD, iD2] = (NeigboursList[iD, iD2]>-1) ?
Math.Sqrt((xCoords[NeigboursList[iD, iD2]] - xCoords[iD]) * (xCoords[NeigboursList[iD, iD2]] -
xCoords[iD]) + (yCoords[NeigboursList[iD, iD2]] - yCoords[iD]) * (yCoords[NeigboursList[iD, iD2]] -
yCoords[iD])) : 0;
           }
        }

        if (Model == "Wave")
        {
           Quadrant = new int[NumData];
           int[] Counter = new int[4];
           for (int iD = 0; iD < NumData; iD++) // For each site
           {
              Quadrant[iD] = (int)TempQuadrant[iD]; // Identify quadrant
              CentroidX[Quadrant[iD] - 1] += xCoords[iD]; // Add up all the x coordinates within the
quadrant
              CentroidY[Quadrant[iD] - 1] += yCoords[iD];
              Counter[Quadrant[iD] - 1]++;
           }

           for (int Q = 0; Q < 4; Q++)// for each quadrant
           {
              CentroidX[Q] /= (double)Counter[Q]; // divide by number of data points to obtain an
average
              CentroidY[Q] /= (double)Counter[Q];
           }
        }
        else if (Model == "Wave2")
        {
           Quadrant = new int[NumData];
           int Counter = 0;
           for (int iD = 0; iD < NumData; iD++) // For each site
           {
              Quadrant[iD] = 1; // Identify quadrant
              CentroidX[0] += xCoords[iD]; // Add up all the x coordinates within the quadrant
              CentroidY[0] += yCoords[iD];
              Counter++;
           }

           CentroidX[0] /= (double)Counter; // divide by number of data points to obtain an average
           CentroidY[0] /= (double)Counter;
        }
     }
```

```
        // Reads in the data from the datafile
        public void ReadInData()
        {
          // Read in the datafile and the data
          DataSet RoadDataFile = DataSet.Open(DataPath + InputDataFile);
          RoadData2004 = RoadDataFile.GetData<double[]>("density04");
          RoadData2007 = RoadDataFile.GetData<double[]>("density07");
          xCoords = RoadDataFile.GetData<double[]>("Longitude");
          yCoords = RoadDataFile.GetData<double[]>("Latitude");
          Location = RoadDataFile.GetData<double[]>("Location");
          int NumData = RoadData2004.Length;
          IsValid = new bool[NumData];
          for (int site = 0; site < NumData; site++) IsValid[site] = true;
          Coded = RoadDataFile.GetData<double[]>("Coded");
          //Coded = new double[NumData];
          //for (int site = 0; site < NumData; site++) Coded[site] = 0;
          if (InputDataFile=="raw_data_Municipio.csv") IsValid[213] = false;
          RoadDataFile.Dispose();
        }

        /// <summary>
        /// Assigns fold numbers to each site
        /// </summary>
        public void AssignFolds()
        {
          int NumData = RoadData2004.Length; // Work out how much data we need
          FoldIndex = new int[NumData]; // stores fold numbers
          double[] RandomNo = new double[NumData]; //random numbers will be put here
          for (int ii = 0; ii < NumData; ii++) // for each site
          {
            Math.DivRem(ii, NFolds, out FoldIndex[ii]); // assign a fold number
            RandomNo[ii] = OurRandomNumberGenerator.NextDouble(); // assign a random number
          }

          Array.Sort(RandomNo, FoldIndex); // sort by random numbers to make random permutation
of folds
        }

        // Sets up the output file with empty containers for the data
        public void SetupOutput()
        {
          DataSet OutputData = DataSet.Open("msds:csv?file=" + DataPath + OutputDataFile +
"&openMode=create");
          OutputData.IsAutocommitEnabled = false;
          string[] DimsArray2 = { "OutputIndex", "Fold" };
          if (!(Model.Contains("Wave")))
          {
            OutputData.Add<double[,]>("r", DimsArray2);
            if (!(Model == "Exponential")) OutputData.Add<double[,]>("K", DimsArray2);
            if (Model.Contains("Dispers")) OutputData.Add<double[,]>("d", DimsArray2);
            if (Model == "Dispersal3") OutputData.Add<double[,]>("Kd", DimsArray2);
```

```
        if (Model == "Dispersal4") OutputData.Add<double[,]>("dk", DimsArray2);
        if (Model.Contains("Dispers")) OutputData.Add<double[,]>("Thresh", DimsArray2);
    }
    else
    {
        if (Model == "Wave")
        {
            for (int par = 1; par < 5; par++)
            {
                OutputData.Add<double[,]>("r"+par.ToString(), DimsArray2);
                //OutputData.Add<double[,]>("K" + par.ToString(), DimsArray2);
                OutputData.Add<double[,]>("Dist" + par.ToString(), DimsArray2);
                OutputData.Add<double[,]>("Angle" + par.ToString(), DimsArray2);
                OutputData.Add<double[,]>("m" + par.ToString(), DimsArray2);
            }
            //OutputData.Add<double[,]>("r", DimsArray2);
            OutputData.Add<double[,]>("K", DimsArray2);
        }
        else if (Model=="Wave2")
        {
            OutputData.Add<double[,]>("r", DimsArray2);
            OutputData.Add<double[,]>("Dist", DimsArray2);
            OutputData.Add<double[,]>("Angle", DimsArray2);
            OutputData.Add<double[,]>("m", DimsArray2);
            OutputData.Add<double[,]>("K", DimsArray2);
        }
    }
    OutputData.Add<double[,]>("theta", DimsArray2);
    OutputData.Add<double[,]>("TL", DimsArray2);
    OutputData.Add<double[,]>("VL", DimsArray2);
    OutputData.Add<double[]>("DIC", "Fold");
    string[] DimsArray2b = { "Interval", "Fold" };
    OutputData.Add<double[,]>("CC", DimsArray2b);
    OutputData.Add<double[,]>("CD", DimsArray2b);


    string[] DimsArray3 = { "Location", "Fold" };
    OutputData.Add<double[,]>("ExamplePredictions", DimsArray3);
    OutputData.Add<double[,]>("PredL95", DimsArray3);
    OutputData.Add<double[,]>("PredMed", DimsArray3);
    OutputData.Add<double[,]>("PredU95", DimsArray3);

    OutputData.Add<double[]>("ProbTL", "Fold");
    OutputData.Add<double[]>("ProbVL", "Fold");

    OutputData.Commit();
}

// Sets up the parameters we want to estimate
public void SetupParametersToEstimate()
{
```

```
        if (!(Model.Contains("Wave")))
        {
          Filzbach.parameter_create("r", 0.0001, 4, 1, 0, 0, 1); // Scales the rate of growth of roads
          if (!(Model == "Exponential")) Filzbach.parameter_create("K", 0.001, 1.0, 0.4, 0, 0, 1); //
Scales the maximum road density
          if (Model.Contains("Dispers")) Filzbach.parameter_create("d", 0.000001, 1.0, 0.0007, 1, 0,
1); // Scales the maximum road density
          if (Model == "Dispersal3") Filzbach.parameter_create("Kd", 0.000001, 1.0, 0.0007, 0, 0, 1);
// Scales the maximum road density
          if (Model == "Dispersal4") Filzbach.parameter_create("dk", -1.0, 1.0, 0.0007, 0, 0, 1); //
Scales the maximum road density
          if (Model.Contains("Dispers")) Filzbach.parameter_create("Thresh", -0.5, 0.5, 0.0, 0, 0, 1);
// Scales the maximum road density
        }
        else
        {
          if (Model == "Wave")
          {
            Filzbach.parameter_create_vector("r", 0.0001, 4, 1, 0, 0, 1, 4);
            //Filzbach.parameter_create_vector("K", 0.001, 1.0, 0.4, 0, 0, 1, 4);
            //Filzbach.parameter_create("r", 0.0001, 4, 1, 0, 0, 1);
            Filzbach.parameter_create("K", 0.001, 1.0, 0.4, 0, 0, 1);
            Filzbach.parameter_create_vector("Dist", 0.00001, 20.0, 0.1, 0, 0, 1, 4);
            Filzbach.parameter_create_vector("Angle", -0.5, 2.2 * Math.PI, 0.1, 0, 0, 1, 4);
            Filzbach.parameter_create_vector("m", 0.0001, 8.0, 0.1, 0, 0, 1, 4);
          }
          else if (Model == "Wave2")
          {
            Filzbach.parameter_create("r", 0.0001, 4, 1, 0, 0, 1);
            Filzbach.parameter_create("K", 0.001, 1.0, 0.4, 0, 0, 1);
            Filzbach.parameter_create("Dist", 0.00001, 20.0, 0.1, 0, 0, 1);
            Filzbach.parameter_create("Angle", -0.5, 2.2 * Math.PI, 0.1, 0, 0, 1);
            Filzbach.parameter_create("m", 0.0001, 8.0, 0.1, 0, 0, 1);
          }
        }
        Filzbach.parameter_create("theta", 0.001, 2, 0.1, 1, 0, 1); // Scales the process error
      }

      // Works out the likelihood of the parameterised model given the data
      public void CalculateLikelihoodFilzbach()
      {
        // First, clear the estimate of likelihood
        Filzbach.set_metr_ltotnew(0.0);

        // Read in the parameter values
        double r_param = 0.0;
        if (!(Model=="Wave")) r_param = Filzbach.parameter_getvalue("r");
        double theta_param = Filzbach.parameter_getvalue("theta");
        double K_param = 0;
        double d_param = 0;
        double dk_param = 0;
```

```
        double Kd_param = 0;
        double Dist = 0.0;
        double Angle = 0.0;
        double m = 0.0;
        double Thresh_param = 0.0;

        if (!(Model.Contains("Wave")))
        {
          if (Model.Contains("Dispers") || Model == "Logistic") K_param =
Filzbach.parameter_getvalue("K");
          if (Model.Contains("Dispers")) d_param = Filzbach.parameter_getvalue("d");
          if (Model == "Dispersal3") Kd_param = Filzbach.parameter_getvalue("Kd");
          if (Model == "Dispersal4") Kd_param = Filzbach.parameter_getvalue("dk");
          if (Model.Contains("Dispers")) Thresh_param = Filzbach.parameter_getvalue("Thresh");
        }

        int NumData = RoadData2004.Length; // Work out how much data we need
        double Prediction = 0.0; // declare a variable to store predictions
        double prob = 0.0; // declare a variable to store the probability value
        double SumL = 0.0; // Declare a variable to hold the log likelihood
        CountZero = 0;

        ImedData1 = new double[NumData];
        ImedData2 = new double[NumData];
        BackupData = new double[NumData];
        for (int iD = 0; iD < NumData; iD++) // For each data point
        {
          BackupData[iD] = RoadData2004[iD];
          ImedData1[iD] = 0.0;
          ImedData2[iD] = 0.0;
        }

        // This could be tidied up but it works
        if (!(Model.Contains("Wave")))
        {
          for (int tt = 0; tt < 3; tt++) // For each data point
          {
            for (int iD = 0; iD < NumData; iD++) // For each data point
            {
              if (Training)
              {
                if (!(FoldIndex[iD] == FoldNumber) && IsValid[iD])
                {
                  if (RoadData2007[iD] > BackupData[iD] && (BackupData[iD] > 0.0))
                  {
                    if (Model == "Exponential")
                    {
                      Prediction = ExpFunction(RoadData2004[iD], r_param, 1.0);
                    }
                    else if (Model == "Logistic")
                    {
```

```
                        Prediction = LogisticFunction(RoadData2004[iD], K_param, r_param, 1.0);
                    }
                    else if (Model == "Dispersal" || Model == "Dispersal1")
                    {
                        Prediction = DispersalFunction(RoadData2004[iD], K_param, r_param,
d_param, Thresh_param, iD, 1.0);
                    }
                    else if (Model == "Dispersal2")
                    {
                        Prediction = DispersalFunction2(RoadData2004[iD], K_param, r_param,
d_param, Thresh_param, iD, 1.0);
                    }
                    else if (Model == "Dispersal3")
                    {
                        Prediction = DispersalFunction3(RoadData2004[iD], K_param, r_param,
d_param, Thresh_param, Kd_param, iD, 1.0);
                    }
                    else if (Model == "Dispersal4")
                    {
                        Prediction = DispersalFunction3(RoadData2004[iD], K_param, r_param,
d_param, Thresh_param, dk_param, iD, 1.0);
                    }
                    ImedData1[iD] = RoadData2004[iD] + Prediction;
                    //prob = Filzbach.normal_density(Math.Log10(RoadData2007[iD] -
RoadData2004[iD]), Math.Log10(Prediction - RoadData2004[iD]), theta_param); // Assume
lognormally distributed process error centred on the predictions
                    //Filzbach.normal_density(Math.Log10(RoadData2007[iD]),
Math.Log10(ImedData[iD]), theta_param); // Assume lognormally distributed process error centred
on the predictions
                    if (tt == 2) prob = Filzbach.normal_density(Math.Log10(RoadData2007[iD] -
BackupData[iD]), Math.Log10(Prediction), theta_param); // Assume lognormally distributed process
error centred on the predictions
                    if (tt == 2) SumL += Math.Log(prob); // Log the probability
                }
                else if (tt == 2) CountZero++;
            }
        }
        else
        {
            if ((FoldIndex[iD] == FoldNumber) && IsValid[iD])
            {
                if (RoadData2007[iD] > BackupData[iD] && (BackupData[iD] > 0.0))
                {
                    if (Model == "Exponential")
                    {
                        Prediction = ExpFunction(RoadData2004[iD], r_param, 1.0);
                    }
                    else if (Model == "Logistic")
                    {
                        Prediction = LogisticFunction(RoadData2004[iD], K_param, r_param, 1.0);
                    }
```

```
                    else if (Model == "Dispersal" || Model == "Dispersal1")
                    {
                      Prediction = DispersalFunction(RoadData2004[iD], K_param, r_param,
d_param, Thresh_param, iD, 1.0);
                    }
                    else if (Model == "Dispersal2")
                    {
                      Prediction = DispersalFunction2(RoadData2004[iD], K_param*(1-Coded[iD]),
r_param, d_param, Thresh_param, iD, 1.0);
                    }
                    else if (Model == "Dispersal3")
                    {
                      Prediction = DispersalFunction3(RoadData2004[iD], K_param, r_param,
d_param, Thresh_param, Kd_param, iD, 1.0);
                    }
                    else if (Model == "Dispersal4")
                    {
                      Prediction = DispersalFunction3(RoadData2004[iD], K_param, r_param,
d_param, Thresh_param, dk_param, iD, 1.0);
                    }
                    ImedData1[iD] = RoadData2004[iD] + Prediction;
                    //Filzbach.normal_density(Math.Log10(RoadData2007[iD]),
Math.Log10(ImedData[iD]), theta_param); // Assume lognormally distributed process error centred
on the predictions
                    if (tt == 2) prob = Filzbach.normal_density(Math.Log10(RoadData2007[iD] -
BackupData[iD]), Math.Log10(Prediction), theta_param); // Assume lognormally distributed process
error centred on the predictions
                    if (tt == 2) SumL += Math.Log(prob); // Log the probability
                  }
                  else if (tt == 2) CountZero++;
                }
              }
            }
          }
          for (int iD = 0; iD < NumData; iD++) RoadData2004[iD] = ImedData1[iD];
        }
        else
        {
          for (int iD = 0; iD < NumData; iD++) // For each data point
          {
            if (Training)
            {
              if (!(FoldIndex[iD] == FoldNumber) && IsValid[iD])
              {
                if (RoadData2004[iD] > 0.0)
                {
                  if (Model=="Wave")
                  {
                    //r_param = Filzbach.parameter_getvalue("r");
                    K_param = Filzbach.parameter_getvalue("K");
                    r_param = Filzbach.parameter_getvalue("r", Quadrant[iD] - 1);
```

```
                    //K_param = Filzbach.parameter_getvalue("K", Quadrant[iD] - 1);
                    Dist = Filzbach.parameter_getvalue("Dist", Quadrant[iD] - 1);
                    Angle = Filzbach.parameter_getvalue("Angle", Quadrant[iD] - 1);
                    m = Filzbach.parameter_getvalue("m", Quadrant[iD] - 1);
                }
                else if (Model == "Wave2")
                {
                    K_param = Filzbach.parameter_getvalue("K");
                    r_param = Filzbach.parameter_getvalue("r");
                    Dist = Filzbach.parameter_getvalue("Dist");
                    Angle = Filzbach.parameter_getvalue("Angle");
                    m = Filzbach.parameter_getvalue("m");
                }

                Prediction = WaveFunction(r_param, K_param, CentroidX[Quadrant[iD] - 1],
CentroidY[Quadrant[iD] - 1], xCoords[iD], yCoords[iD], Dist, Angle, 0);
                prob = Filzbach.normal_density(Math.Log10(RoadData2004[iD]),
Math.Log10(Prediction), theta_param); // Assume lognormally distributed process error centred on
the predictions
                SumL += Math.Log(prob); // Log the probability
                Prediction = WaveFunction(r_param, K_param, CentroidX[Quadrant[iD] - 1],
CentroidY[Quadrant[iD] - 1], xCoords[iD], yCoords[iD], Dist, Angle, m);
                ImedData2[iD] = Prediction;
                prob = Filzbach.normal_density(Math.Log10(RoadData2007[iD]),
Math.Log10(Prediction), theta_param); // Assume lognormally distributed process error centred on
the predictions
                SumL += Math.Log(prob); // Log the probability
                ImedData1[iD] = Prediction;

                //Prediction = WaveFunction(r_param, K_param, CentroidX[Quadrant[iD] - 1],
CentroidY[Quadrant[iD] - 1], xCoords[iD], yCoords[iD], Dist, Angle, m)-WaveFunction(r_param,
K_param, CentroidX[Quadrant[iD] - 1], CentroidY[Quadrant[iD] - 1], xCoords[iD], yCoords[iD], Dist,
Angle, 0);
                //prob = Filzbach.normal_density(Math.Log10(RoadData2007[iD] -
RoadData2004[iD]), Math.Log10(Prediction), theta_param);
                //SumL += Math.Log(prob); // Log the probability

              }
              else CountZero++;
            }
          }
          else
          {
            if ((FoldIndex[iD] == FoldNumber) && IsValid[iD])
            {
              if (RoadData2004[iD] > 0)
              {
                if (Model=="Wave")
                {
                    //r_param = Filzbach.parameter_getvalue("r");
                    K_param = Filzbach.parameter_getvalue("K");
```

```
                    r_param = Filzbach.parameter_getvalue("r", Quadrant[iD] - 1);
                    //K_param = Filzbach.parameter_getvalue("K", Quadrant[iD] - 1);
                    Dist = Filzbach.parameter_getvalue("Dist", Quadrant[iD] - 1);
                    Angle = Filzbach.parameter_getvalue("Angle", Quadrant[iD] - 1);
                    m = Filzbach.parameter_getvalue("m", Quadrant[iD] - 1);
                  }
                  else if (Model == "Wave2")
                  {
                    K_param = Filzbach.parameter_getvalue("K");
                    r_param = Filzbach.parameter_getvalue("r");
                    Dist = Filzbach.parameter_getvalue("Dist");
                    Angle = Filzbach.parameter_getvalue("Angle");
                    m = Filzbach.parameter_getvalue("m");
                  }

                  Prediction = WaveFunction(r_param, K_param, CentroidX[Quadrant[iD] - 1],
CentroidY[Quadrant[iD] - 1], xCoords[iD], yCoords[iD], Dist, Angle, 0);
                  prob = Filzbach.normal_density(Math.Log10(RoadData2004[iD]),
Math.Log10(Prediction), theta_param); // Assume lognormally distributed process error centred on
the predictions
                  SumL += Math.Log(prob); // Log the probability
                  Prediction = WaveFunction(r_param, K_param, CentroidX[Quadrant[iD] - 1],
CentroidY[Quadrant[iD] - 1], xCoords[iD], yCoords[iD], Dist, Angle, m);
                  ImedData2[iD] = Prediction;
                  prob = Filzbach.normal_density(Math.Log10(RoadData2007[iD]),
Math.Log10(Prediction), theta_param); // Assume lognormally distributed process error centred on
the predictions
                  SumL += Math.Log(prob); // Log the probability
                  ImedData1[iD] = Prediction;
                }
                else CountZero++;
              }
            }
          }
        }
        CurrentLikelihood = SumL;
        Filzbach.inc_metr_ltotnew(SumL); // increment the estimate of likelihood.
        for (int iD = 0; iD < NumData; iD++) RoadData2004[iD] = BackupData[iD];
      }

      // Outputs the data to a datafile
      public void OutputDataToFileFolds()
      {
        DataSet OutputData = DataSet.Open("msds:csv?file=" + DataPath + OutputDataFile);
        int NumData = RoadData2004.Length; // Work out how much data we need

        int numParameterSets = 0;// counts the number of outputs from the Bayes List
        while (Filzbach.params_from_bayes_list(0, numParameterSets) == 0) numParameterSets++;
        double[] rValues = new double[numParameterSets]; // These are stores for the parameters
that will be output although they are not all output - it depends on the model being fitted
        double[] KValues = new double[numParameterSets];
```

```
        double[] KdValues = new double[numParameterSets];
        double[] dkValues = new double[numParameterSets];
        double[] dValues = new double[numParameterSets];
        double[] ThreshValues = new double[numParameterSets];
        double[] thetaValues = new double[numParameterSets];
        double[,] AngleiValues = new double[numParameterSets,4];
        double[,] DistiValues = new double[numParameterSets,4];
        double[,] miValues = new double[numParameterSets,4];
        double[] AngleValues = new double[numParameterSets];
        double[] DistValues = new double[numParameterSets];
        double[] mValues = new double[numParameterSets];
        double[,] riValues = new double[numParameterSets,4];
        double[,] KiValues = new double[numParameterSets,4];
        int[] idValues = new int[numParameterSets];
        double[] ValidationLikelihoods = new double[numParameterSets]; // Saves the evaluation
likelihoods associated with each set of parameter values
        double[] TrainingLikelihoods = new double[numParameterSets]; // Saves the training
likelihoods associated with each set of parameter values
        double[] RawValidationLikelihoods = new double[numParameterSets]; // Saves the
evaluation likelihoods associated with each set of parameter values
        double[] RawTrainingLikelihoods = new double[numParameterSets]; // Saves the training
likelihoods associated with each set of parameter values
        double LogProbsModelTL = 0.0;
        double LogProbsModelVL = 0.0;


        int[] CountFolds = new int[NFolds]; // used to work out average likelihoods
        double[] DensityData = new double[NumData];
        int CountValid = 0;
        for (int F = 0; F < NumData; F++) CountFolds[FoldIndex[F]]++;

        for (int Dat = 0; Dat < NumData; Dat++)
        {
          if (IsValid[Dat])
          {
            //if (!(Model.Contains("Wave")))
            //{
              if (RoadData2007[Dat] > RoadData2004[Dat] && (RoadData2004[Dat] > 0.0))
              {
                DensityData[Dat] = Math.Log10(RoadData2007[Dat] - RoadData2004[Dat] + 1.0);
                CountValid++;
              }
            /*}
            else
            {
              if (RoadData2004[Dat] > 0.0)
              {
                DensityData[Dat] = Math.Log10(RoadData2007[Dat] + 1.0);
                CountValid++;
              }
            }*/
```

```
    }
  }


  // These are some performance metrics
  double MeanLikelihood = 0.0;
  double[] CC = new double[numParameterSets]; // stores the correlation coefficicent
  double[] CD = new double[numParameterSets]; // stores the coefficient of determination
  int[] SiteIndex = new int[NumData];
  double[] PredictionList = new double[NumData];
  double[] PredictionList2 = new double[NumData];
  double[] PredictionListRecord = new double[NumData];
  double[,] PredictionCatalogue = new double[NumData, numParameterSets];
  double[] PredictionL95 = new double[NumData];
  double[] PredictionMed = new double[NumData];
  double[] PredictionU95 = new double[NumData];

  Training = false;
  this.CalculateLikelihoodFilzbach();
  int CountData = CountFolds[FoldNumber] - CountZero; // CC, CD calculation step
  double[] VPred = new double[CountData];// CC, CD calculation step
  double[] VObs = new double[CountData];// CC, CD calculation step

  double MaxLLTL = -1000000;
  double MaxLLVL = -1000000;

  for (int iDP = 0; iDP < numParameterSets; iDP++) // For each data point
  {
    // Record all of the sampled parameter values
    Filzbach.params_from_bayes_list(0, iDP);
    thetaValues[iDP] = Filzbach.cv("theta");
    if (!(Model.Contains("Wave")))
    {
      rValues[iDP] = Filzbach.cv("r");
      if (Model == "Dispersal3") KdValues[iDP] = Filzbach.cv("Kd");
      if (Model == "Dispersal4") dkValues[iDP] = Filzbach.cv("dk");
      if (Model.Contains("Dispers")) ThreshValues[iDP] = Filzbach.cv("Thresh");
      if (Model.Contains("Dispers")) dValues[iDP] = Filzbach.cv("d");
      if (!(Model == "Exponential")) KValues[iDP] = Filzbach.cv("K");
    }
    else
    {
      if (Model == "Wave")
      {
        KValues[iDP] = Filzbach.cv("K");
        //rValues[iDP] = Filzbach.cv("r");
        for (int Q = 0; Q < 4; Q++)
        {
          riValues[iDP, Q] = Filzbach.cv("r", Q);
          //KiValues[iDP, Q] = Filzbach.cv("K", Q);
          AngleiValues[iDP, Q] = Filzbach.cv("Angle", Q);
```

```
                DistiValues[iDP, Q] = Filzbach.cv("Dist", Q);
                miValues[iDP, Q] = Filzbach.cv("m", Q);
            }
        }
        else if (Model == "Wave2")
        {
            KValues[iDP] = Filzbach.cv("K");
            rValues[iDP] = Filzbach.cv("r");
            AngleValues[iDP] = Filzbach.cv("Angle");
            DistValues[iDP] = Filzbach.cv("Dist");
            mValues[iDP] = Filzbach.cv("m");
        }
    }

    idValues[iDP] = iDP;
    PredictionList = new double[NumData];
    PredictionList2 = new double[NumData];

    // Add up the likelihoods for the evaluation data
    Training = false;
    this.CalculateLikelihoodFilzbach();
    RawValidationLikelihoods[iDP] = CurrentLikelihood;
    ValidationLikelihoods[iDP] = CurrentLikelihood / ((double)CountFolds[FoldNumber] -
CountZero); // normalie the estimaton of likelihood by the number of data points used for
assessment
    if (Model.Contains("Wave")) ValidationLikelihoods[iDP] /= 2.0; // The wave model training
step combines 2004 and 2007 data so we additionally divide by 2
    for (int iD = 0; iD < NumData; iD++) if (FoldNumber == FoldIndex[iD]) PredictionList[iD] =
ImedData1[iD];
    for (int iD = 0; iD < NumData; iD++) if (FoldNumber == FoldIndex[iD]) PredictionList2[iD] =
ImedData2[iD];

    MaxLLVL = (CurrentLikelihood > MaxLLVL) ? CurrentLikelihood : MaxLLVL;

    // do the same for training data, calculating DIC too
    Training = true;
    this.CalculateLikelihoodFilzbach();
    MeanLikelihood += -2 * CurrentLikelihood; //DIC Calculation step
    RawTrainingLikelihoods[iDP] = CurrentLikelihood;
    TrainingLikelihoods[iDP] = CurrentLikelihood / ((double)NumData - CountZero -
((double)CountFolds[FoldNumber]));
    if (Model.Contains("Wave")) TrainingLikelihoods[iDP] /= 2.0; // The wave model training
step combines 2004 and 2007 data so we additionally divide by 2
    for (int iD = 0; iD < NumData; iD++) if (!(FoldNumber == FoldIndex[iD])) PredictionList[iD] =
ImedData1[iD];
    for (int iD = 0; iD < NumData; iD++) if (FoldNumber == FoldIndex[iD]) PredictionList2[iD] =
ImedData2[iD];

    MaxLLTL = (CurrentLikelihood > MaxLLTL) ? CurrentLikelihood : MaxLLTL;

    // Here we record average predictions for each data point
```

```
            int Counter = 0;
            for (int iD = 0; iD < NumData; iD++)
            {
              // Save the predictions so we can do stuff with them like calc prediction, CC and CD
intervals
              if (!(Model.Contains("Wave")))
              {
                PredictionList[iD] -= RoadData2004[iD];
                PredictionCatalogue[iD, iDP] = Math.Log10(PredictionList[iD] + 1.0);
                PredictionListRecord[iD] += PredictionList[iD];
                if ((FoldIndex[iD] == FoldNumber) && (RoadData2007[iD] > RoadData2004[iD]) &&
(RoadData2004[iD] > 0.0) && IsValid[iD])
                {
                  VPred[Counter] = Math.Log10(PredictionList[iD] + 1.0);
                  VObs[Counter] = Math.Log10(RoadData2007[iD] - RoadData2004[iD] + 1.0);
                  Counter++;
                }
              }
              else
              {
                PredictionCatalogue[iD, iDP] = PredictionList[iD];
                PredictionListRecord[iD] += PredictionList[iD];
                if ((FoldIndex[iD] == FoldNumber) && (RoadData2004[iD] > 0.0) && IsValid[iD])
                {
                  //VPred[Counter] =  Math.Log10(PredictionList[iD] + 1.0);
                  //VObs[Counter] = Math.Log10(RoadData2007[iD] + 1.0);
                  VPred[Counter] = Math.Log10(PredictionList[iD] - PredictionList2[iD] + 1.0);
                  VObs[Counter] = Math.Log10(RoadData2007[iD] - RoadData2004[iD] + 1.0);
                  Counter++;
                }
              }
            }
            CC[iDP] = PearsonProductMomentCorrelationCoefficient(VObs, VPred);
            CD[iDP] += CoefficientOfDetermination(VObs, VPred);
          }

          for (int iDP = 0; iDP < numParameterSets; iDP++) // For each data point
          {
            LogProbsModelTL += Math.Exp(RawTrainingLikelihoods[iDP] - MaxLLTL);
            LogProbsModelVL += Math.Exp(RawValidationLikelihoods[iDP] - MaxLLVL);
          }
          LogProbsModelTL = MaxLLTL + Math.Log(LogProbsModelTL);
          LogProbsModelVL = MaxLLVL + Math.Log(LogProbsModelVL);

          for (int iD = 0; iD < NumData; iD++) PredictionListRecord[iD] /= (double)numParameterSets;
          double[] CCInt = Intervals95(CC, false);
          double[] CDInt = Intervals95(CD, false);

          for (int iD = 0; iD < NumData; iD++)
          {
            double[] preds = new double[numParameterSets];
```

```
            for (int iDP = 0; iDP < numParameterSets; iDP++) // For each data point
            {
               preds[iDP] = (PredictionCatalogue[iD, iDP]<0.0) ? 0.0: PredictionCatalogue[iD, iDP];
            }
            double[] intervals = Intervals95(preds, false);
            PredictionL95[iD] = intervals[0];
            PredictionMed[iD] = intervals[1];
            PredictionU95[iD] = intervals[2];
         }


         MeanLikelihood /= (double)numParameterSets;//DIC Calculation step
         Filzbach.params_set_to_posterior_mean(); // set the parameters to their posterior mean
         Training = true;//DIC Calculation step
         this.CalculateLikelihoodFilzbach();//DIC Calculation step
         double MeanExpectationLikelihood = -2 * CurrentLikelihood;//DIC Calculation step
         double EffectiveNumParameters = MeanLikelihood - MeanExpectationLikelihood;//DIC
Calculation step
         Training = true;

         OutputData.IsAutocommitEnabled = false;
         if (FoldNumber == 0) OutputData.Add<int[]>("OutputIndex", idValues, "OutputIndex");
         OutputData.Append<double[]>("theta", thetaValues, "Fold");
         OutputData.Append<double[]>("TL", TrainingLikelihoods, "Fold");
         OutputData.Append<double[]>("VL", ValidationLikelihoods, "Fold");
         if (FoldNumber == 0) OutputData.Add<int[]>("FoldAssignment", FoldIndex, "Location");
         //if (FoldNumber == 0) for (int iD = 0; iD < NumData; iD++) SiteIndex[iD] = iD;
         if (FoldNumber == 0) OutputData.Add<double[]>("Location", Location, "Location");
         if (FoldNumber == 0) OutputData.Add<double[]>("DeltaDensity", DensityData, "Location");
         if (FoldNumber == 0) OutputData.Add<bool[]>("isValid", IsValid, "Location");
         if (FoldNumber == 0 && (Model== "Wave")) OutputData.Add<int[]>("Quadrant", Quadrant,
"Location");
         if (FoldNumber == 0 && (Model.Contains("Wave")))
OutputData.Add<double[]>("RoadData2007", RoadData2007, "Location");
         OutputData.Append<double[]>("ExamplePredictions", PredictionListRecord, "Fold");
         OutputData.Append<double[]>("PredL95", PredictionL95, "Fold");
         OutputData.Append<double[]>("PredMed", PredictionMed, "Fold");
         OutputData.Append<double[]>("PredU95", PredictionU95, "Fold");

         if (!(Model.Contains("Wave")))
         {
            OutputData.Append<double[]>("r", rValues, "Fold");
            if (!(Model == "Exponential")) OutputData.Append<double[]>("K", KValues, "Fold");
            if ( Model == "Dispersal3") OutputData.Append<double[]>("Kd", KdValues, "Fold");
            if (Model == "Dispersal4") OutputData.Append<double[]>("dk", KdValues, "Fold");
            if (Model.Contains("Dispers")) OutputData.Append<double[]>("Thresh", ThreshValues,
"Fold");
            if (Model.Contains("Dispers")) OutputData.Append<double[]>("d", dValues, "Fold");
         }
         else
         {
```

```
            if (Model == "Wave")
            {
                for (int par = 1; par < 5; par++)
                {
                    double[] AValues = new double[numParameterSets];
                    double[] DValues = new double[numParameterSets];
                    double[] miVal = new double[numParameterSets];
                    for (int iDP = 0; iDP < numParameterSets; iDP++) // For each data point
                    {
                        rValues[iDP] = riValues[iDP, par - 1];
                        //KValues[iDP] = KiValues[iDP, par-1];
                        AValues[iDP] = AngleiValues[iDP, par - 1];
                        DValues[iDP] = DistiValues[iDP, par - 1];
                        miVal[iDP] = miValues[iDP, par - 1];
                    }
                    //OutputData.Append<double[]>("r", rValues, "Fold");
                    if (par == 1) OutputData.Append<double[]>("K", KValues, "Fold");
                    OutputData.Append<double[]>("r" + par.ToString(), rValues, "Fold");
                    //OutputData.Append<double[]>("K" + par.ToString(), KValues, "Fold");
                    OutputData.Append<double[]>("Angle" + par.ToString(), AValues, "Fold");
                    OutputData.Append<double[]>("Dist" + par.ToString(), DValues, "Fold");
                    OutputData.Append<double[]>("m" + par.ToString(), miVal, "Fold");
                }
            }
            else if (Model == "Wave2")
            {
                //OutputData.Append<double[]>("r", rValues, "Fold");
                OutputData.Append<double[]>("K", KValues, "Fold");
                OutputData.Append<double[]>("r", rValues, "Fold");
                OutputData.Append<double[]>("Angle", AngleValues, "Fold");
                OutputData.Append<double[]>("Dist", DistValues, "Fold");
                OutputData.Append<double[]>("m" , mValues, "Fold");
            }
        }

        OutputData.Append<double>("ProbTL", LogProbsModelTL);
        OutputData.Append<double>("ProbVL", LogProbsModelVL);

        OutputData.Append<double>("DIC", EffectiveNumParameters + MeanLikelihood,
"Fold");//DIC Calculation step
        OutputData.Append<double[]>("CC", CCInt, "Fold");//DIC Calculation step
        OutputData.Append<double[]>("CD", CDInt, "Fold");//DIC Calculation step
        OutputData.Commit();
    }

    /// <summary>
    /// Our model predicting road density growth as an exponential function
    /// </summary>
    /// <param name="R0">Initial road density at 2004</param>
    /// <param name="r">Per capita rate of change of road density</param>
    /// <returns>The prediction of road density at 2007</returns>
```

```
    public double ExpFunction(double R0, double r, double t)
    {
      return r * t; // predict the new road density using the exponential equation
    }

    /// <summary>
    /// Our model predicting road density growth as a logistic function
    /// </summary>
    /// <param name="R0">Initial road density at 2004</param>
    /// <param name="K">Maximum road density</param>
    /// <param name="r">Per capita rate of change of road density</param>
    /// <returns>The prediction of road density at 2007</returns>
    public double LogisticFunction(double R0, double K, double r, double t)
    {
      //double Expr3 = Math.Exp(r * t); // This simply saves time  - we only need to do exp(r*3)
once
      //return R0 * K * Expr3 / (K + R0 * (Expr3 - 1.0)); // predict the new road density using the
logistic equation
      double Expr3 = R0 * r * (1.0 - R0 / K) * t;
      return (Expr3 > 0.0) ? Expr3 : 0.0;
    }

    /// <summary>
    /// Returns the result of a saturating function model
    /// </summary>
    /// <param name="Alpha">The maximum value of the function</param>
    /// <param name="Kappa">The half saturation constant</param>
    /// <param name="N">The value to assess the function at</param>
    /// <returns>The value of the function</returns>
    public double SaturatingFunction(double Alpha, double Kappa, double N)
    {
      return Alpha * N / (Kappa + N); // predict the new road density using the logistic equation
    }

    /// <summary>
    /// Our model predicting road density growth as a combination of logistic growth and
neighbourhood effects
    /// where neighbourhood effects are simple linear functions, though these may incorporate
Euclidean distance
    /// </summary>
    /// <param name="R0">Initial road density at 2004</param>
    /// <param name="K">Maximum road density</param>
    /// <param name="r">Per capita rate of change of road density</param>
    /// <param name="d">The dispersal parameter</param>
    /// <param name="iD">The site ID number</param>
    /// <param name="t">The toime over which to conduct the integration</param>
    /// <returns>The prediction of road density at 2007</returns>
    public double DispersalFunction(double R0, double K, double r, double d, double Thresh, int iD,
double t)
    {
      double InternalGrowth = LogisticFunction(R0, K, r, t);
```

```
        double NewDensity = 0.0;
        for (int Neighbour = 0; Neighbour < NumNeighbours[iD]; Neighbour++)
        {
          if (NeigboursList[iD, Neighbour] > -1)
          {
            if (IsValid[NeigboursList[iD, Neighbour]] && Coded[NeigboursList[iD, Neighbour]] <
CodedThresh && (RoadData2004[NeigboursList[iD, Neighbour]] > 0.0))
            {
              if (DisperalSubmodel == 1) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? (RoadData2004[NeigboursList[iD, Neighbour]] -
RoadData2004[iD]-Thresh) : 0.0; // Movement rate is proportional to difference only
              else if (DisperalSubmodel == 2) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? (RoadData2004[NeigboursList[iD, Neighbour]] -
RoadData2004[iD] - Thresh) / EuclideanDistList[iD, Neighbour] : 0.0; // Movement rate is
proportional to gradient of differences
              else if (DisperalSubmodel == 3) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
(RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - Thresh) : 0.0; // Per road
movement rate is proportional to differences
              else if (DisperalSubmodel == 4) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
(RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - Thresh) / EuclideanDistList[iD,
Neighbour] : 0.0; // per road movement rate is proportional to gradient of differences
              /*if (DisperalSubmodel == 1) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]); // Movement rate is proportional to difference only
              else if (DisperalSubmodel == 2) NewDensity +=  (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) / EuclideanDistList[iD, Neighbour]; // Movement rate is
proportional to gradient of differences
              else if (DisperalSubmodel == 3) NewDensity +=  RoadData2004[NeigboursList[iD,
Neighbour]] * (RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD]); // Per road
movement rate is proportional to differences
              else if (DisperalSubmodel == 4) NewDensity += RoadData2004[NeigboursList[iD,
Neighbour]] * (RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD]) /
EuclideanDistList[iD, Neighbour]; // per road movement rate is proportional to gradient of
differences*/
            }
          }
        }
        double Prediction = LogisticFunction(RoadData2004[iD], K, r, t) + NewDensity * d * t;
        return Prediction;
      }

      /// <summary>
      /// Our model predicting road density growth as a combination of logistic growth and
neighbourhood effects
      /// where neighbourhood effects are simple exponential functions, though these may
incorporate Euclidean distance
      /// </summary>
      /// <param name="R0">Initial road density at 2004</param>
      /// <param name="K">Maximum road density</param>
      /// <param name="r">Per capita rate of change of road density</param>
```

```
        /// <param name="d">Scales the overall rate of dispersal</param>
        /// <param name="iD">The site ID number</param>
        /// <param name="t">The time over which to conduct the integration</param>
        /// <returns>The prediction of road density at 2007</returns>
        public double DispersalFunction2(double R0, double K, double r, double d, double Thresh, int
iD, double t)
        {
            double InternalGrowth = LogisticFunction(R0, K, r, t);
            double NewDensity = 0.0;
            for (int Neighbour = 0; Neighbour < NumNeighbours[iD]; Neighbour++)
            {
                if (NeigboursList[iD, Neighbour] > -1)
                {
                    if (IsValid[NeigboursList[iD, Neighbour]] && Coded[NeigboursList[iD, Neighbour]] <
CodedThresh && (RoadData2004[NeigboursList[iD, Neighbour]] > 0.0))
                    {
                        if (DisperalSubmodel == 1) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? Math.Exp((RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD] - Thresh)) : 0.0; // per road movement rate is exponential function
of positve difference
                        else if (DisperalSubmodel == 2) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? Math.Exp((RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD] - Thresh) / EuclideanDistList[iD, Neighbour]) : 0.0; // per road
movement rate is exponential function of positve difference
                        else if (DisperalSubmodel == 3) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
Math.Exp(((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - Thresh))) : 0.0; // per
road movement rate is exponential function of positve difference
                        else if (DisperalSubmodel == 4) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
Math.Exp(((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - Thresh) /
EuclideanDistList[iD, Neighbour])) : 0.0; // per road movement rate is exponential function of
positve difference*/
                        /*if (DisperalSubmodel == 1) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? Math.Exp((RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD] - 0)) : 0.0; // per road movement rate is exponential function of
positve difference
                        else if (DisperalSubmodel == 2) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? Math.Exp((RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD] - 0) / EuclideanDistList[iD, Neighbour]) : 0.0; // per road movement
rate is exponential function of positve difference
                        else if (DisperalSubmodel == 3) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
Math.Exp(((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - 0))) : 0.0; // per road
movement rate is exponential function of positve difference
                        else if (DisperalSubmodel == 4) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
Math.Exp(((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - 0) /
EuclideanDistList[iD, Neighbour])) : 0.0; // per road movement rate is exponential function of
positve difference*/
```

```
                    /*if (DisperalSubmodel == 1) NewDensity += 
Math.Exp((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD])); // per road 
movement rate is exponential function of positve difference
                    else if (DisperalSubmodel == 2) NewDensity += (RoadData2004[NeigboursList[iD, 
Neighbour]] - RoadData2004[iD]) > 0.0 ? RoadData2004[NeigboursList[iD, Neighbour]] * 
Math.Exp(((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD]))) : 0.0; // per road 
movement rate is exponential function of positve difference
                    else if (DisperalSubmodel == 3) NewDensity += 
Math.Exp((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD]) / EuclideanDistList[iD, 
Neighbour]); // per road movement rate is exponential function of positve difference

                    else if (DisperalSubmodel == 4) NewDensity += (RoadData2004[NeigboursList[iD, 
Neighbour]] - RoadData2004[iD]) > 0.0 ? RoadData2004[NeigboursList[iD, Neighbour]] * 
Math.Exp(((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD]) / 
EuclideanDistList[iD, Neighbour])) : 0.0; // per road movement rate is exponential function of 
positve difference*/
                }
            }
        }
        double Prediction = LogisticFunction(RoadData2004[iD], K, r, t) + NewDensity * d * t;
        return Prediction;
    }



    /// <summary>
    /// Our model predicting road density growth as a combination of logistic growth and 
neighbourhood effects
    /// where neighbourhood effects are simple exponential functions, though these may 
incorporate Euclidean distance
    /// </summary>
    /// <param name="R0">Initial road density at 2004</param>
    /// <param name="K">Maximum road density</param>
    /// <param name="r">Per capita rate of change of road density</param>
    /// <param name="d">Scales the overall rate of dispersal</param>
    /// <param name="iD">The site ID number</param>
    /// <param name="t">The time over which to conduct the integration</param>
    /// <returns>The prediction of road density at 2007</returns>
    public double DispersalFunction3(double R0, double K, double r, double d, double Thresh, 
double Kd, int iD, double t)
    {
        double InternalGrowth = LogisticFunction(R0, K, r, t);
        double NewDensity = 0.0;
        for (int Neighbour = 0; Neighbour < NumNeighbours[iD]; Neighbour++)
        {
            if (NeigboursList[iD, Neighbour] > -1)
            {
                if (IsValid[NeigboursList[iD, Neighbour]] && Coded[NeigboursList[iD, Neighbour]] < 
CodedThresh && (RoadData2004[NeigboursList[iD, Neighbour]] > 0.0))
                {
```

```
                    if (DisperalSubmodel == 1) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? SaturatingFunction(1.0, Kd,
RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - Thresh) : 0.0; // per road
movement rate is exponential function of positve difference
                    else if (DisperalSubmodel == 2) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? SaturatingFunction(1.0, Kd,
(RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - Thresh) / EuclideanDistList[iD,
Neighbour]) : 0.0; // per road movement rate is exponential function of positve difference
                    else if (DisperalSubmodel == 3) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
SaturatingFunction(1.0, Kd, RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] -
Thresh) : 0.0; // per road movement rate is exponential function of positve difference
                    else if (DisperalSubmodel == 4) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
SaturatingFunction(1.0, Kd, (RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] -
Thresh) / EuclideanDistList[iD, Neighbour]) : 0.0; // per road movement rate is exponential function
of positve difference
                }
            }
        }
        double Prediction = LogisticFunction(RoadData2004[iD], K, r, t) + NewDensity * d * t;
        return Prediction;
    }

    public double DispersalFunction4(double R0, double K, double r, double d, double Thresh,
double dk, int iD, double t)
    {
        double InternalGrowth = LogisticFunction(R0, K, r, t);
        double NewDensity = 0.0;
        for (int Neighbour = 0; Neighbour < NumNeighbours[iD]; Neighbour++)
        {
            if (NeigboursList[iD, Neighbour] > -1)
            {
                if (IsValid[NeigboursList[iD, Neighbour]] && Coded[NeigboursList[iD, Neighbour]] <
CodedThresh && (RoadData2004[NeigboursList[iD, Neighbour]] > 0.0))
                {
                    if (DisperalSubmodel == 1) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? Math.Exp((RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD] - Thresh+dk)) : 0.0; // per road movement rate is exponential
function of positve difference
                    else if (DisperalSubmodel == 2) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? Math.Exp((RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD] - Thresh + dk) / EuclideanDistList[iD, Neighbour]) : 0.0; // per road
movement rate is exponential function of positve difference
                    else if (DisperalSubmodel == 3) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
Math.Exp(((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - Thresh + dk))) : 0.0;
// per road movement rate is exponential function of positve difference
                    else if (DisperalSubmodel == 4) NewDensity += (RoadData2004[NeigboursList[iD,
Neighbour]] - RoadData2004[iD]) > Thresh ? RoadData2004[NeigboursList[iD, Neighbour]] *
Math.Exp(((RoadData2004[NeigboursList[iD, Neighbour]] - RoadData2004[iD] - Thresh + dk) /
```

```
EuclideanDistList[iD, Neighbour])) : 0.0; // per road movement rate is exponential function of
positve difference*/
                }
            }
        }
        double Prediction = LogisticFunction(RoadData2004[iD], K, r, t) + NewDensity * d * t;
        return Prediction;
    }
    /// <summary>
    /// Calculates the road density as if it is a traveling wave over the land surface.
    /// </summary>
    /// <param name="r">Rate of change of road density with distance</param>
    /// <param name="K">Maximum road density</param>
    /// <param name="cX">X coordinate for centroid of quadrant</param>
    /// <param name="cY">Y coordinate for centroid of quadrant</param>
    /// <param name="sX">X coordinate for municipio</param>
    /// <param name="sY">Y coordinate for municipio</param>
    /// <param name="Distance">Estimated euclidean distance from centre of wave to centroid of
quadrant in 2004</param>
    /// <param name="Angle">Estimated angle of wave</param>
    /// <param name="m">Estimated movement distance of wave by 2007</param>
    /// <returns>Prediction of density at location</returns>
    public double WaveFunction(double r, double K, double cX, double cY, double sX, double sY,
double Distance, double Angle, double m)
    {
        double CentX = cX + Distance * Math.Sin(Angle); // X coordinate of centre of wave
        double CentY = cY + Distance * Math.Cos(Angle); // Y coordinate of centre of wave
        double Newy = sX * Math.Sin(Angle) + sY * Math.Cos(Angle) + cY; //Y coodrinate of municipio
along angle of projection
        double NewCy = CentX * Math.Sin(Angle) + CentY * Math.Cos(Angle) + cY; // Y coordinate of
centre of wave along angle of projection
        double Dist = Newy - NewCy; // Distance from minicipio centre to centre of wave

        double Prediction = K * (1.0 / (1.0 + Math.Exp(-r * (Dist+m))));

        return Prediction;
    }

    /// <summary>
    /// Returns the Pearson Product Moment Correlation Coefficient -  one way to measure how
much
    /// variance in the data is explained by the predictions of a model.
    /// </summary>
    /// <param name="Observations">The observed data</param>
    /// <param name="Predictions">The predicted data</param>
    /// <returns>A single value of R-squared - the squared correlation coefficient</returns>
    public static double PearsonProductMomentCorrelationCoefficient(double[] Observations,
double[] Predictions)
    {
        double ObservationSumOfSquares = 0.0;
        double PredictionSumOfSquares = 0.0;
```

```
            double CovariationSumOfSquares = 0.0;
            double MeanObservations = 0.0;
            double MeanPredictions = 0.0;
            int numObservations = Observations.Length;
            double FracObservations = 1.0 / (double)numObservations;

            foreach (double Observation in Observations) MeanObservations += Observation *
FracObservations;
            foreach (double Prediction in Predictions) MeanPredictions += Prediction * FracObservations;

            for (int ii = 0; ii < numObservations; ii++)
            {
                CovariationSumOfSquares += (Observations[ii] - MeanObservations) * (Predictions[ii] -
MeanPredictions);
                ObservationSumOfSquares += (Observations[ii] - MeanObservations) * (Observations[ii] -
MeanObservations);
                PredictionSumOfSquares += (Predictions[ii] - MeanPredictions) * (Predictions[ii] -
MeanPredictions);
            }

            if ((PredictionSumOfSquares > 0) && (ObservationSumOfSquares > 0))
            {
                return (CovariationSumOfSquares / (Math.Sqrt(ObservationSumOfSquares) *
Math.Sqrt(PredictionSumOfSquares)));
            }
            else return 0;
        }

        /// <summary>
        /// Calculates the general "Coefficient of Determination".
        /// Definition:
        /// </summary>
        /// <param name="Observations">The observed data</param>
        /// <param name="Predictions">The predicted data</param>
        /// <returns></returns>
        public static double CoefficientOfDetermination(double[] Observations, double[] Predictions)
        {
            double TotalSumOfSquares = 0.0;
            double ErrorSumOfSquares = 0.0;
            double MeanObservations = 0.0;
            int numObservations = Observations.Length;

            foreach (double Observation in Observations) MeanObservations += Observation;
            MeanObservations /= (double)numObservations;

            for (int ii = 0; ii < numObservations; ii++)
            {
                TotalSumOfSquares += (Observations[ii] - MeanObservations) * (Observations[ii] -
MeanObservations);
                ErrorSumOfSquares += (Observations[ii] - Predictions[ii]) * (Observations[ii] -
Predictions[ii]);
```

```
        }

        return 1.0 - (ErrorSumOfSquares / TotalSumOfSquares);
    }

    /// <summary>
    /// Returns 2.5th and 97.5th percentiles and the median for a data array passed to it
    /// </summary>
    /// <param name="Data">The data to draw the intervals for</param>
    /// <param name="LogDataFirst">"True" means log transoform the data before
processing</param>
    /// <returns>a tuple containing 1. the lowe 2.5th percentile, 2. the median and 3. the upper
97.5th percentile of the data</returns>
    public static double[] Intervals95(double[] Data, bool LogDataFirst)
    {
        int nSamples = Data.Length;

        if (LogDataFirst) for (int ii = 0; ii < nSamples; ii++) Data[ii] = Math.Log(Data[ii]);

        if (nSamples > 3)
        {
            int L95 = (int)Math.Ceiling(0.025 * (double)nSamples);
            int U95 = nSamples - L95;
            int Med = (int)Math.Round(0.5 * (double)nSamples);

            Array.Sort(Data);

            double L95Data = Data[L95];
            double MedData = Data[Med];
            double U95Data = Data[U95];

            double[] Results = { L95Data, MedData, U95Data };
            return Results;
        }
        else if (nSamples > 0)
        {
            int Med = (int)Math.Round(0.5 * (double)nSamples);

            Array.Sort(Data);

            double MedData = Data[Med];

            double[] Results = { 0, MedData, 0 };
            return Results;
        }
        else
        {
            double[] Results = { 0, 0, 0 };
            return Results;
        }
    }
```

```
    public static void Simulate(string Model, int DispersalSubmodel, string NameOfAnalysis)
    {
      //Run this to simulate from 2004 to 2104

      //Load parameters into memory

      //Set up grid of outputs

      //output data

      Console.WriteLine("Program has started"); // Flag to the user that the program initialised
successfully.

      // Points to the function used to assess likelihood
      // It is unlikely that you will need to change these two lines
      // but to use Filzbach using C# we need to create a Program "object"
      ModelCode FittingProgram = new ModelCode();
      FittingProgram.OurRandomNumberGenerator = new Random(1);
      FittingProgram.Model = Model;
      FittingProgram.DisperalSubmodel = DispersalSubmodel;
      FittingProgram.RunID = NameOfAnalysis + FittingProgram.Model +
FittingProgram.DisperalSubmodel.ToString();
      FittingProgram.OutputDataFile = FittingProgram.RunID + "OutputData.csv";


      // This function reads in the data from the datafile
      FittingProgram.ReadInData();
      DataSet RoadDataFile = DataSet.Open(FittingProgram.DataPath +
FittingProgram.InputDataFile);
      FittingProgram.Coded = RoadDataFile.GetData<double[]>("Coded");

      FittingProgram.AssignNeighbours();

      FittingProgram.PerformSimulations();
    }

    public void PerformSimulations()
    {
      //DataSet ParametersFile = DataSet.Open(DataPath + OutputDataFile);
      DataSet ParametersFile = DataSet.Open(DataPath + "Grid100Dispersal21OutputData.csv");
      //Here is where to adjust to get the right datafile.

      int SimYears = 100;
      int NData = RoadData2004.Length;

      int NumData = RoadData2004.Length; // Work out how much data we need

      int numParameterSets = 0;// counts the number of outputs from the Bayes List
      double[,] rValues = ParametersFile.GetData<double[,]>("r");
```

```
        numParameterSets = rValues.GetLength(0);
        double[,] KValues = new double[numParameterSets, NFolds];
        double[,] KdValues = new double[numParameterSets, NFolds];
        double[,] dValues = new double[numParameterSets, NFolds];
        double[,] ThreshValues = new double[numParameterSets, NFolds];
        double[,] DistValues = new double[numParameterSets, NFolds];
        double[,] AngleValues = new double[numParameterSets, NFolds];
        double[,] mValues = new double[numParameterSets, NFolds];
        DataSet LocationDataFile = DataSet.Open(DataPath + InputDataFile);
        double[] Location =LocationDataFile.GetData<double[]>("Location");

        if (!(Model == "Exponential")) KValues = ParametersFile.GetData<double[,]>("K");
        if (Model.Contains("Dispers")) dValues = ParametersFile.GetData<double[,]>("d");
        if (Model == "Dispersal3") KdValues = ParametersFile.GetData<double[,]>("Kd");
        if (Model.Contains("Dispers")) ThreshValues = ParametersFile.GetData<double[,]>("Thresh");
        if (Model == "Wave2")
        {
          DistValues = ParametersFile.GetData<double[,]>("Dist");
          AngleValues = ParametersFile.GetData<double[,]>("Angle");
          mValues = ParametersFile.GetData<double[,]>("m");
        }


        double[,] DensityPredictionsU95 = new double[NData, SimYears];
        double[,] DensityPredictionsL95 = new double[NData, SimYears];
        double[,] DensityPredictionsMed = new double[NData, SimYears];
        double[] NextState = new double[NData];
        double[] CurrentState = new double[NData];
        double[] RoadData2004Backup = new double[NData];
        short[, ,] TempStore = new short[numParameterSets, NData, SimYears];

        for (int F = 0; F < 10; F++)
        {
          for (int iDP = 0; iDP < numParameterSets; iDP++) // For each data point
          {
            for (int iD = 0; iD < NumData; iD++)
            {
              NextState[iD] = RoadData2004[iD];
              RoadData2004Backup[iD] = RoadData2004[iD];
              Coded[iD] = 0;
            }

            for (int Y = 0; Y < 100; Y++)
            {
              for (int iD = 0; iD < NumData; iD++)
              {
                  if (Model == "Exponential")
                  {
                    NextState[iD] = RoadData2004[iD] + ExpFunction(RoadData2004[iD],
rValues[iDP, F], 1.0);
                  }
```

```
                    else if (Model == "Logistic")
                    {
                      NextState[iD] = RoadData2004[iD] + LogisticFunction(RoadData2004[iD],
KValues[iDP, F], rValues[iDP, F], 1.0);
                    }
                    else if (Model == "Dispersal" || Model == "Dispersal1")
                    {
                      NextState[iD] = RoadData2004[iD] + DispersalFunction(RoadData2004[iD],
KValues[iDP, F], rValues[iDP, F], dValues[iDP, F], ThreshValues[iDP, F], iD, 1.0);
                    }
                    else if (Model == "Dispersal2")
                    {
                      NextState[iD] = RoadData2004[iD] + DispersalFunction2(RoadData2004[iD],
KValues[iDP, F], rValues[iDP, F], dValues[iDP, F], ThreshValues[iDP, F], iD, 1.0);
                    }
                    else if (Model == "Dispersal3")
                    {
                      NextState[iD] = RoadData2004[iD] + DispersalFunction3(RoadData2004[iD],
KValues[iDP, F], rValues[iDP, F], dValues[iDP, F], ThreshValues[iDP, F], KdValues[iDP, F], iD, 1.0);
                    }
                    else if (Model == "Wave2")
                    {
                      NextState[iD] = WaveFunction(rValues[iDP, F], KValues[iDP, F],
CentroidX[Quadrant[iD] - 1], CentroidY[Quadrant[iD] - 1], xCoords[iD], yCoords[iD], DistValues[iDP,
F], AngleValues[iDP, F], (double)Y * (mValues[iDP, F] / 3.0));
                    }
                    if (Coded[iD] >= CodedThresh) NextState[iD] = RoadData2004[iD];
                    TempStore[iDP, iD, Y] = (short)(NextState[iD]*10000);
                  }
                for (int iD = 0; iD < NumData; iD++) RoadData2004[iD] = NextState[iD];
              }

              for (int iD = 0; iD < NumData; iD++) RoadData2004[iD] = RoadData2004Backup[iD];
            }

            for (int Y = 0; Y < 100; Y++)
            {
              for (int iD = 0; iD < NumData; iD++)
              {
                double[] TempDataStore = new double[numParameterSets];
                for (int iDP = 0; iDP < numParameterSets; iDP++) // For each data point
                {
                  TempDataStore[iDP] = ((double)TempStore[iDP, iD, Y])/10000.0;
                }
                double[] limits = Intervals95(TempDataStore, false);
                DensityPredictionsL95[iD, Y] += limits[0];
                DensityPredictionsMed[iD, Y] += limits[1];
                DensityPredictionsU95[iD, Y] += limits[2];

              }
            }
```

```
        }

        for (int Y = 0; Y < 100; Y++)
        {
          for (int iD = 0; iD < NumData; iD++)
          {
            DensityPredictionsL95[iD, Y] /= (double)NFolds;
            DensityPredictionsMed[iD, Y] /= (double)NFolds;
            DensityPredictionsU95[iD, Y] /= (double)NFolds;

          }
        }


        DataSet SimDataOutputFile = DataSet.Open("msds:csv?file=" + DataPath + "SimulationData"
+ AnalysisName + Model + DisperalSubmodel.ToString()+ ".csv&openMode=create");
        //SimDataOutputFile.Add<double[]>("Longitude", xCoords, "Point");
        //SimDataOutputFile.Add<double[]>("Latitude", yCoords, "Point");
        SimDataOutputFile.Add<double[]>("Location", Location, "Location");
        SimDataOutputFile.Add<double[]>("Longitude", xCoords, "Location");
        SimDataOutputFile.Add<double[]>("Latitude", yCoords, "Location");
        string[] dimensions = {"Location","Year"};
        SimDataOutputFile.Add<double[,]>("OutputL95", DensityPredictionsL95, dimensions);
        SimDataOutputFile.Add<double[,]>("OutputMed", DensityPredictionsMed, dimensions);
        SimDataOutputFile.Add<double[,]>("OutputU95", DensityPredictionsU95, dimensions);
        SimDataOutputFile.Commit();
      }

      public static void SimulateGrid(string Model, int DispersalSubmodel, string NameOfAnalysis)
      {
        //Run this to simulate from 2004 to 2104

        //Load parameters into memory

        //Set up grid of outputs

        //output data

        Console.WriteLine("Program has started"); // Flag to the user that the program initialised
successfully.

        // Points to the function used to assess likelihood
        // It is unlikely that you will need to change these two lines
        // but to use Filzbach using C# we need to create a Program "object"
        ModelCode FittingProgram = new ModelCode();
        FittingProgram.OurRandomNumberGenerator = new Random(1);
        FittingProgram.Model = Model;
        FittingProgram.DisperalSubmodel = DispersalSubmodel;
        FittingProgram.RunID = NameOfAnalysis + FittingProgram.Model +
FittingProgram.DisperalSubmodel.ToString();
        FittingProgram.OutputDataFile = FittingProgram.RunID + "OutputData.csv";
```

```
            // This function reads in the data from the datafile
            FittingProgram.ReadInData();
            DataSet RoadDataFile = DataSet.Open(FittingProgram.DataPath +
FittingProgram.InputDataFile);
            FittingProgram.Coded = RoadDataFile.GetData<double[]>("Coded");

            //FittingProgram.CoarsenRoadGrid();

            FittingProgram.AssignNeighbours();

            FittingProgram.PerformSimulationsGrid();
        }

        public void CoarsenRoadGrid()
        {
            DataSet RoadData = DataSet.Open(DataPath + "RoadDensityMap.nc");
            DataSet NewRoadData = DataSet.Open(DataPath + "NewRoadDensityMap.nc");
            double[,] RoadsGrid04 = RoadData.GetData<double[,]>("RoadsResamp");
            double[,] Protected = RoadData.GetData<double[,]>("Protected");
            double[,] Rivers = RoadData.GetData<double[,]>("Rivers");
            int lats = RoadsGrid04.GetLength(0);
            int lons = RoadsGrid04.GetLength(1);

            int NumNewLats = lats / 10;
            int NumNewLons = lons / 10;

            double[,] NewRoadsGrid04 = new double[NumNewLats, NumNewLons];
            double[,] NewProtected = new double[NumNewLats, NumNewLons];
            double[,] NewRivers = new double[NumNewLats, NumNewLons];

            for (int xx = 0; xx < NumNewLats; xx++)
            {
                for (int yy = 0; yy < NumNewLons; yy++)
                {
                    int latstart = xx * 10;
                    int lonstart = yy * 10;
                    for (int xx2 = latstart; xx2 < latstart + 11; xx2++)
                    {
                        for (int yy2 = lonstart; yy2 < lonstart + 11; yy2++)
                        {
                            NewRoadsGrid04[xx,yy]+=RoadsGrid04[xx2,yy2];
                            NewRivers[xx,yy] += (Rivers[xx2,yy2]<1.0) ? 0.0 : 1.0;
                            NewProtected[xx,yy] += (Protected[xx2,yy2]<1.0) ? 0.0 : 1.0;
                        }
                    }
                    NewRoadsGrid04[xx, yy] /= 121.0;
                    NewRivers[xx, yy] = (NewRivers[xx, yy] < 50) ? 0 : 1;
                    NewProtected[xx, yy] = (NewProtected[xx, yy] < 50) ? 0 : 1;
                }
```

```
    }

    string[] dims1 = { "y", "x" };
    NewRoadData.Add<double[,]>("RoadGrid04", NewRoadsGrid04, dims1);
    NewRoadData.Add<double[,]>("Protected", NewProtected, dims1);
    NewRoadData.Add<double[,]>("Rivers", NewRivers, dims1);
    NewRoadData.Dispose();
}


public void PerformSimulationsGrid()
{
    //DataSet ParametersFile = DataSet.Open(DataPath + OutputDataFile);
    DataSet ParametersFile = DataSet.Open(DataPath + "Grid100Dispersal21OutputData.csv");
    //Here is where to adjust to get the right datafile.
    /*DataSet Protect = DataSet.Open(DataPath + "protect.asc");
    DataSet Rivers = DataSet.Open(DataPath + "rivers.asc");
    DataSet Roads04 = DataSet.Open(DataPath + "roads04.csv?InferDims=true");
    DataSet Roads07 = DataSet.Open(DataPath + "roads07.csv?InferDims=true");
    var protectdata = Protect.GetData<double[,]>("protect");
    var riversdata = Rivers.GetData<double[,]>("rivers");
    var RoadsID04 = Roads04.GetData<double[]>("number");
    var RoadsID07 = Roads04.GetData<double[]>("number");
    var RoadsLength04 = Roads04.GetData<double[]>("length04");
    var RoadsLength07 = Roads07.GetData<double[]>("length07");
    int lats = protectdata.GetLength(0);
    int lons = protectdata.GetLength(1);
    int samp04 = RoadsID04.Length;
    int samp07 = RoadsID07.Length;*/
    DataSet RoadData = DataSet.Open(DataPath + "NewRoadDensityMap.nc");

    /*double[,] RoadsGrid04 = new double[lats, lons ];
    double[,] RoadsGrid07 = new double[lats, lons];
    double[,] RoadsDiff = new double[lats, lons];
    double[,] RoadsResamp = new double[lats, lons];
    double[,] DataDummy1 = new double[lats, lons];
    double[,] DataDummy2 = new double[lats, lons];
    int buff = 50;
    int avger=(buff*2+1)*(buff*2+1);

    for (int ii = 0; ii < samp04; ii++)
    {
        int xcd = (int)RoadsID04[ii] % lats;
        int ycd = (int)RoadsID04[ii] / lats;
        RoadsGrid04[xcd, ycd] = RoadsLength04[ii] / 1000.0;
    }

    for (int xx = 0; xx < lats; xx++)
    {
        for (int yy = 0; yy < lons; yy++)
        {
```

```
        for (int xw = xx-buff; xw<(xx+buff); xw++)
        {
          if ((xw>-1) && (xw<lats))
          {
            for (int yw = yy - buff; yw < (yy + buff); yw++)
            {
              if ((yw > -1) && (yw < lons))
              {
                RoadsResamp[xx, yy] += RoadsGrid04[xw, yw];
              }
            }
          }
        }
      }
    }


    for (int xx = 0; xx < lats; xx++)
    {
      for (int yy = 0; yy < lons; yy++)
      {
        RoadsResamp[xx,yy]/= (double)avger;
        DataDummy1[xx, yy] = protectdata[lats-xx-1, yy];
        DataDummy2[xx, yy] = riversdata[lats-xx-1, yy];
      }
    }



    for (int ii = 0; ii < samp07; ii++)
    {
      int xcd = (int)RoadsID07[ii] % lats;
      int ycd = (int)RoadsID07[ii] / lats;
      RoadsGrid07[xcd, ycd] = RoadsLength07[ii] / 1000.0;
      RoadsDiff[xcd, ycd] = RoadsGrid07[xcd, ycd] - RoadsGrid04[xcd, ycd];
    }*/

    /*DataSet CreateOutput = DataSet.Open(DataPath + "tester.nc");
    string[] dims1 = {"y","x"};
    CreateOutput.Add<double[,]>("RoadGrid04", RoadsGrid04,dims1);
    CreateOutput.Add<double[,]>("RoadGrid07", RoadsGrid07, dims1);
    CreateOutput.Add<double[,]>("RoadDiffs", RoadsDiff, dims1);
    CreateOutput.Add<double[,]>("Protected", DataDummy1, dims1);
    CreateOutput.Add<double[,]>("Rivers", DataDummy2, dims1);
    CreateOutput.Add<double[,]>("RoadsResamp", RoadsResamp, dims1);
    CreateOutput.Dispose();*/

    double[,] RoadsGrid04 = RoadData.GetData<double[,]>("RoadGrid04");
    double[,] Protected = RoadData.GetData<double[,]>("Protected");
    double[,] Rivers = RoadData.GetData<double[,]>("Rivers");
    int lats = RoadsGrid04.GetLength(0);
```

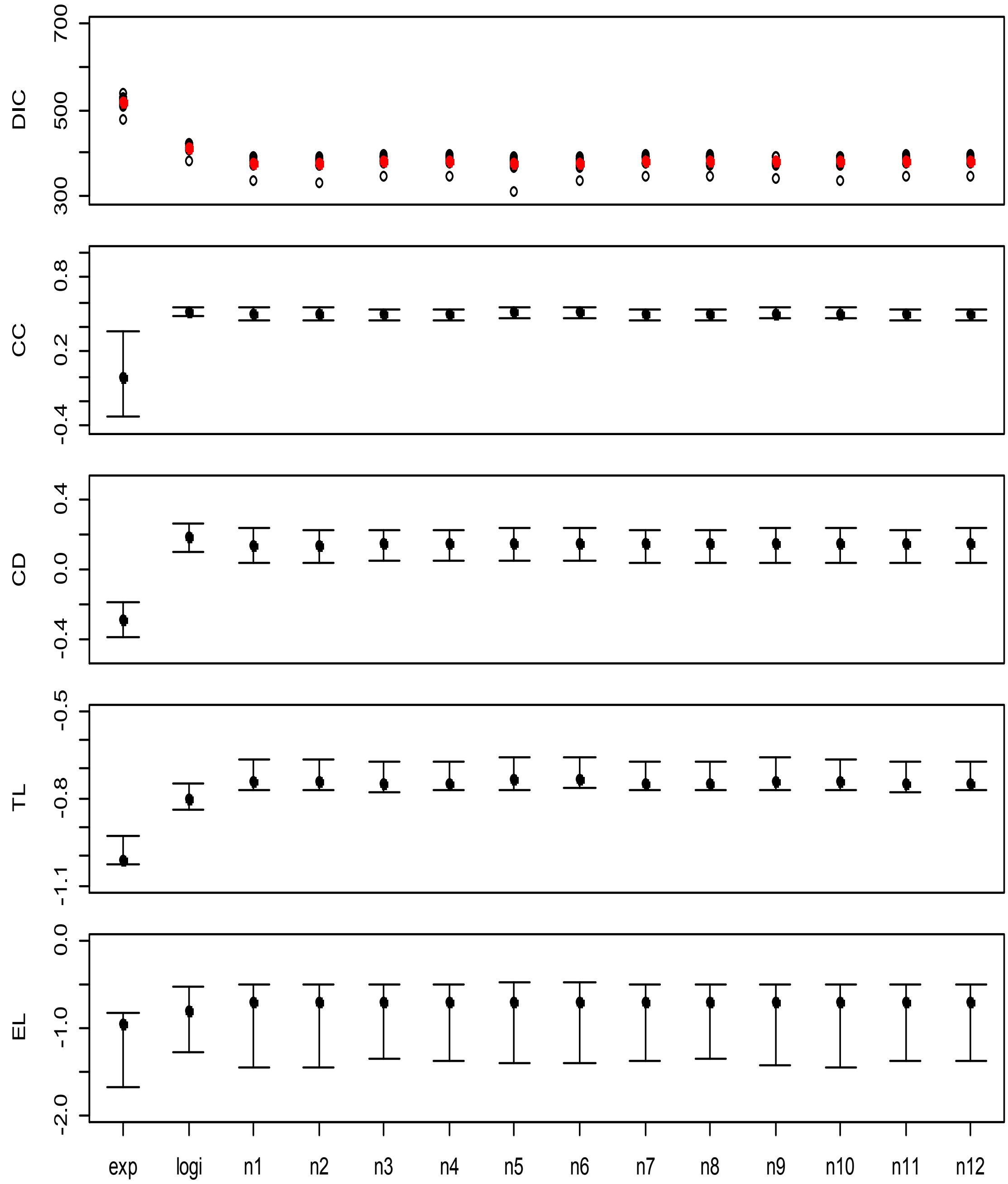
```
        int lons = RoadsGrid04.GetLength(1);
        double[] latitudes = new double[lats];
        double[] longitudes = new double[lons];

        for (int ii = 0; ii < lats; ii++) latitudes[ii] = -17.538 + (double)ii * (0.083333);
        for (int ii = 0; ii < lons; ii++) longitudes[ii] = -74.676 + (double)ii * (0.083333);

        double dt = 0.01;
        double Ddtdx2 = 0.001 * dt / (0.083333 * 0.083333);

        int numParameterSets = 0;// counts the number of outputs from the Bayes List
        double[,] rValues = ParametersFile.GetData<double[,]>("r");
        numParameterSets = rValues.GetLength(0);
        double[,] KValues = new double[numParameterSets, NFolds];
        double[,] KdValues = new double[numParameterSets, NFolds];
        double[,] dValues = new double[numParameterSets, NFolds];
        double[,] ThreshValues = new double[numParameterSets, NFolds];
        double[,] DistValues = new double[numParameterSets, NFolds];
        double[,] AngleValues = new double[numParameterSets, NFolds];
        double[,] mValues = new double[numParameterSets, NFolds];

        if (!(Model == "Exponential")) KValues = ParametersFile.GetData<double[,]>("K");
        if (Model.Contains("Dispers")) dValues = ParametersFile.GetData<double[,]>("d");
        if (Model == "Dispersal3") KdValues = ParametersFile.GetData<double[,]>("Kd");
        if (Model.Contains("Dispers")) ThreshValues = ParametersFile.GetData<double[,]>("Thresh");
        if (Model == "Wave2")
        {
          DistValues = ParametersFile.GetData<double[,]>("Dist");
          AngleValues = ParametersFile.GetData<double[,]>("Angle");
          mValues = ParametersFile.GetData<double[,]>("m");
        }

        //double[, ,] DensityPredictionsSD = new double[lats, lons, 10];
        double[, ,] DensityPredictionsMean = new double[lats, lons, 10];
        double[,] CurrentState = new double[lats, lons];
        double[,] NextState = new double[lats, lons];
        double NEffect = 0.0;
        int lam1 = lats - 1;
        int lom1 = lons - 1;
        DataSet SimDataOutputFile = DataSet.Open(DataPath +
"SimulationDataTest.nc?openMode=create&rollbackEnabled=false");
        string[] dims2 = { "latitude", "longitude" };
        SimDataOutputFile.Add<double[,]>("DensityPredictionsMean0", RoadsGrid04, dims2);

        for (int F = 0; F < 1; F++)
        {
          for (int iDP = 0; iDP < 1; iDP++) // For each data point
          {
            for (int xx = 0; xx < lats; xx++) for (int yy = 0; yy < lons; yy++) CurrentState[xx, yy] =
RoadsGrid04[xx, yy];
```

```
            for (int Y = 1; Y < 61; Y++)
            {
                for (int spacer = 0; spacer < 100; spacer++)
                {
                    //for (int xx = 0; xx < lats; xx++)
                    for (int xx = 0; xx < lats; xx++)
                    {
                        //for (int yy = 0; yy < lons; yy++)
                        for (int yy = 0; yy < lons; yy++)
                        {
                            if ((Protected[xx, yy] < 1.0) && (Rivers[xx, yy] < 1.0))
                            {
                                NEffect = 0.0;
                                for (int xn = xx - 1; xn < xx + 2; xn++)
                                {
                                    if ((xn > -1) && (xn < lats))
                                    {
                                        for (int yn = yy - 1; yn < yy + 2; yn++)
                                        {
                                            if ((yn > -1) && (yn < lons))
                                            {
                                                if (!((xx == xn) && (yy == yn)))
                                                {
                                                    if ((Protected[xn, yn] < 1.0) && (Rivers[xn, yn] < 1.0))
                                                    {
                                                        // NEffect += (CurrentState[xn, yn] - CurrentState[xx, yy]) > 1e-6
? (CurrentState[xn, yn] - CurrentState[xx, yy]) : 0 ; // per road movement rate is exponential
function of positve difference
                                                        NEffect += (CurrentState[xn, yn] - CurrentState[xx, yy]) >
ThreshValues[iDP, F] ? (CurrentState[xn, yn] - CurrentState[xx, yy]) : 0; // per road movement rate is
exponential function of positve difference
                                                    }
                                                }
                                            }
                                        }
                                    }
                                }
                                // NextState[xx, yy] = CurrentState[xx, yy] +
(LogisticFunction(CurrentState[xx, yy], KValues[iDP, F], rValues[iDP, F], dt) + dt *(0.081/rValues[iDP,
F]) * NEffect);
                                NextState[xx, yy] = CurrentState[xx, yy] + (LogisticFunction(CurrentState[xx,
yy], KValues[iDP, F], rValues[iDP, F], dt) + Ddtdx2 * NEffect);
                            }
                            else NextState[xx, yy] = CurrentState[xx, yy];
                        }
                    }
                    for (int xx = 0; xx < lats; xx++) for (int yy = 0; yy < lons; yy++) CurrentState[xx, yy] =
NextState[xx, yy];
                }
                //if ((Y%10==0)&&(Y>0)) for (int xx = 0; xx < lats; xx++) for (int yy = 0; yy < lons; yy++)
DensityPredictionsMean[xx,yy,Ybox] = NextState[xx,yy];
```

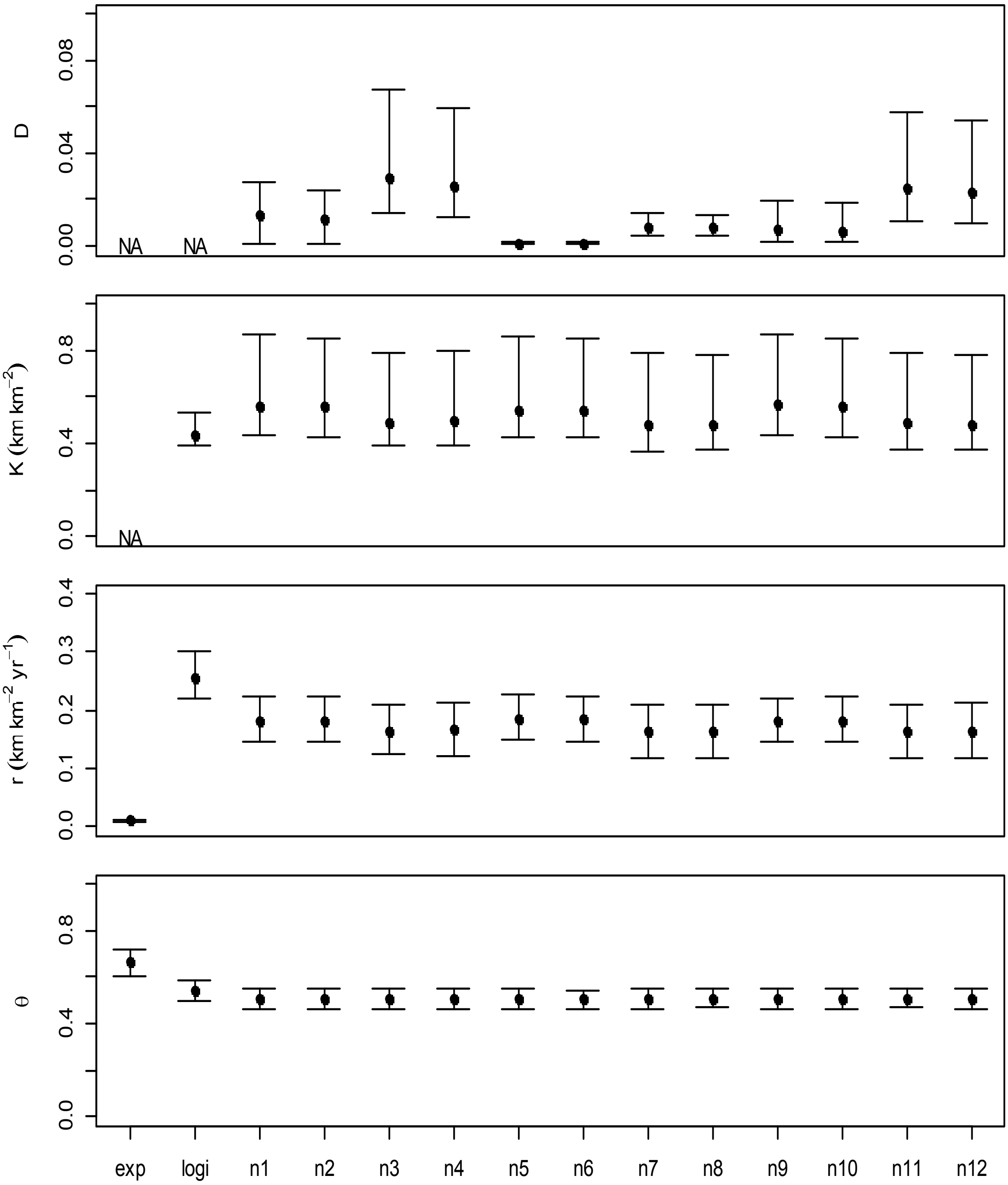
```
                if ((Y == 5) || (Y == 10) || (Y == 25) || (Y == 60))
SimDataOutputFile.Add<double[,]>("DensityPredictionsMean" + Y.ToString(), NextState, dims2);
                Console.WriteLine(Y);
              }
            }
          }

          SimDataOutputFile.Add<double[]>("latitude", latitudes, "latitude");
          SimDataOutputFile.Add<double[]>("longitude", longitudes, "longitude");

          //string[] dims2 = { "x", "y"};
          //SimDataOutputFile.Add<double[ ,]>("DensityPredictionsMean", CurrentState, dims2);
          //SimDataOutputFile.Commit();

          //SimDataOutputFile.Add<double[, ,]>("DensityPredictionsMean", DensityPredictionsMean,
dims3);
          //SimDataOutputFile.Add<double[]>("Longitude", xCoords, "Point");
          //SimDataOutputFile.Add<double[]>("Latitude", yCoords, "Point");
          // SimDataOutputFile.Add<double[]>("Location", Location, "Location");
          //SimDataOutputFile.Add<double[]>("Longitude", xCoords, "Location");
          //SimDataOutputFile.Add<double[]>("Latitude", yCoords, "Location");
          SimDataOutputFile.Commit();
          /*
          string[] dimensions = { "Location", "Year" };
          SimDataOutputFile.Add<double[,]>("OutputL95", DensityPredictionsL95, dimensions);
          SimDataOutputFile.Add<double[,]>("OutputMed", DensityPredictionsMed, dimensions);
          SimDataOutputFile.Add<double[,]>("OutputU95", DensityPredictionsU95, dimensions);
          SimDataOutputFile.Commit();*/
        }
    }
}
```

Appendix B: **Additional figure for Chapter 6**

a)



Appendix B. a) Goodness of fit measures for logistic (logi), exponential (exp) and 4 neighbourhood effects models (equations 4-7 with Exponential functional forms as defined in Chapter 5). Mean parameter values and 95% confidence intervals are displayed for all goodness of fit measures except DIC, for which mean DIC (dark circle) and 10 DICs (grey circles) from each of the 10 fold parameter estimations are displayed. The exponential model performs worst for all measures. DIC=Deviance information criterion, CC= coefficient of correlation, CD= coefficient of determination, TL= training likelihood, EL= evaluation likelihood b)



Appendix B .b) Estimated parameters for logistic, exponential and four neighbourhood effects models (equations 4-7 with Exponential functional forms as defined in section 2.3) at the grid100 scale. Mean parameter values and 95% confidence intervals are displayed. D= magnitude of neighbourhood effect (units differ depending on formulation see section 2.3), K= maximum road density (km km^{-2}), r=maximum road growth rate (km km^{-2} yr^{-1}), τ=road density threshold difference (between neighbours) at which neighbourhood effects become apparent (km km^{-2}), theta = estimated variance in the observations about the model predictions.

Appendix C: **Model code for Chapter 7**

Model code for Chapter 6

// Written in Python for use with ArcGIS
// With Thanks to David Orme

```
# ---------------------------------------------------------------------------
# process_mod_2.1.py
# ---------------------------------------------------------------------------

##  NOTE THAT SMOOTHING OF THE REVENUE SURFACE IS DONE INTERNALLY

## CHANGES
## 2.1 - complete rewrite. The algorithm now searches the local area around
##      roads out to a distance determinded by the density increase and selects
##      a local high value patch.
##    - Unused road length can be recycled within a year.
##    - Patches are grown at the road end to optimise revenue from that roadhead
##    - Masks used to reduce processing time on cost distance runs. Well. They
##      would be if it was remotely obvious how to set the SA environment programatically in 9.3
##    - needs input test data to be complete and not to have seriously blocky values - affects
##      concession size if there are NAs and big blocks of equal area data. Some low value noise
##      works for cost.
## 2.0 - pick target regions by % area not % of max profit.
##    - switched from queens move connections in region definition
##      to rook move
##    - extracted road distance for each potential region to target
##      realistic annual concessions
##    - removed reference to location_poly, originally used to set
##      bounds of analysis but the cost layer is more appropriate
##    - changed setup of parameters to handle working with arc toolbox
##      which passes numeric values as strings.
##    - removed defaults for input files

############################
## Import system modules ##
############################

import sys, string, os, arcgisscripting
import numpy as np    # vectors and arrays
from dbfpy import dbf # csv handling

############################
## ArcGIS initialisation ##
############################

# Create the ArcGIS 9.3 Geoprocessor object
gp = arcgisscripting.create(9.3)
# Set the necessary product code
gp.SetProduct("ArcInfo")
# Check out the spatial analyst license
gp.CheckOutExtension("spatial")
```

```
############################
## Arguments and defaults ##
############################

# testing default string
# D:\Sadia_processModel\FinalTestInputs\revenue.img
D:\Sadia_processModel\FinalTestInputs\scaleCostLow.img
D:\Sadia_processModel\FinalTestInputs\231_65_2000_proj.shp
D:\Sadia_processModel\FinalTestInputs\roadDensity_m_m2.txt
D:\Sadia_processModel\FinalTestInputs\ConcessionSizes.txt Logfile.txt
D:\Sadia_processModel\Scratch2 10 2000 0.25 0.2 0.0001 true

# input layers, change  and concession size table and log file
Revenue_layer = sys.argv[1]

Cost_Layer = sys.argv[2]

Roads = sys.argv[3]

changeTable = sys.argv[4]

concessionTable = sys.argv[5]

logFile = sys.argv[6]

# extract and set the scratch workspace
scratchWorkspace = sys.argv[7]
gp.workspace = scratchWorkspace

# number of years to simulate
numYears = sys.argv[8]
numYears = 10 if numYears == '#' else int(numYears)

# area around roads considered to be worked out
# and also the minimum road distance to be built
roadBuffer = sys.argv[9]
roadBuffer = 2000 if roadBuffer == '#' else float(roadBuffer)

# relative cost of skidder track extraction versus road building.
skidderTrack = sys.argv[10]
skidderTrack = 0.25 if skidderTrack == '#' else float(skidderTrack)

# road growth fuzz
roadRange = sys.argv[11]
roadRange = 0.2 if roadRange == '#' else float(roadRange)

# density range
densityRange = sys.argv[12]
densityRange = 0.1 if densityRange == '#' else float(densityRange)
```

```
# Run with verbose reporting of progress?
verbose = sys.argv[13]
if verbose == '#':
   verbose = True
else:  # handle possible input values.
   boolStr = {"true": True, "t": True, "false": False, "f": False}
   verbose = boolStr.get(verbose.lower())


################################
## Simple utility functions ##
################################

# road length from polyline shapefile

def RoadLength(shpFile):
   # This is 9.3 specific
   # open the feature class as a search cursor
   length=0
   rows = gp.SearchCursor(shpFile)
   row = rows.Next()

   # Enter while loop for each feature/row
   while row:
      feature = row.shape
      length += feature.length
      row = rows.Next()

   del row, rows #delete row object variable
   return length

# count number of non noData cells to get road cells
def RoadCellCount(raster):
  rasTable = gp.SearchCursor(raster)
  val = 0
  rasRow = rasTable.Next()
  while rasRow:
     val += rasRow.GetValue('Count')
     rasRow = rasTable.Next()

  del rasTable, rasRow
  return val

def ReadDBF(path, filename):
  # a)  open file
  fullpath = os.path.join(path, filename)
  in_db = dbf.Dbf(fullpath)
  # b)  load the data rows and close
  fn = in_db.fieldNames
  n = in_db.recordCount
  data = [None] * n
```

```
  for row in range(n): data[row] = in_db[row]
  in_db.close()
  # c)  transpose list of rows into list of numpy arrays
  transpose = [list(c) for c in zip(*data)]
  namedData = dict(zip(fn, [np.array(v) for v in transpose]))
  del data, fn,transpose, in_db
  return(namedData)

########################################
## Load the road density change table ##
########################################

# NB - minimal error checking at present
# Open the file and read two tab-delimited columns:
# - current density, change in density to next year
f = open(changeTable)
dat = f.readlines()
f.close() # Close the file when done.

# drop the header lines
dat = dat[1:]

# first, strip off the carriage returns
datSplit = [x.strip() for x in dat]

# now split the data up
datSplit = [x.split('\t') for x in datSplit]

# extract two variables from the data and turn into numpy arrays
density = np.array([float(x[0]) for x in datSplit])
changedens = np.array([float(x[1]) for x in datSplit])

####################################
## Load the concession size table ##
####################################

# NB - minimal error checking at present
# Open the file and read a single column of sizes
f = open(concessionTable)
dat = f.readlines()
f.close() # Close the file when done.

datSplit = [x.strip() for x in dat]
concessionSizes = np.array([float(x) for x in datSplit])

########################
## Setup the log file ##
########################

# write the logfile in the scratch workspace
os.chdir(scratchWorkspace)
```

```
log = open(logFile, 'w')
# write the parameters to the first line
[log.write(x + '\t') for x in sys.argv[1:]]
log.write('\n')
# write headers
headers = ("Year\tArea\tTotalRoadLength\tTargetDensChange\t" +
        "TargetTotalRoadLength\tbuiltRoadLength\t" +
        "RoadCost" + "\t" + "ConcessionSize" + "\t" +
        "TotalRevenue" + '\t' +  "ExtractCost" + '\t' +  "RCRatio" + "\n")

log.write(headers)
log.flush()

##########################################
## Spatial extent and scale of analysis ##
##########################################

# find the  extent and resolution of the revenue layer as the area of analysis
desc = gp.describe(Revenue_layer)
rastCols = desc.width
rastRows = desc.height
rastNCells = rastRows * rastCols
cellSize = desc.meanCellHeight
rastUnits = desc.spatialreference.linearunitname
Area = rastNCells * cellSize * cellSize

# set the geoprocessor extent and cell size to that of the revenue layer
gp.extent = desc.extent
gp.cellSize = cellSize
del desc

##########################################
## Initial layer conversions and copies ##
##########################################

# Convert the roads file to a binary raster
gp.PolylineToRaster_conversion(Roads, "FID", "road_temp", "MAXIMUM_LENGTH", "NONE",
cellSize)
gp.SingleOutputMapAlgebra_sa("con([road_temp] >= 0,1)", "road_rast")
gp.Delete_management("road_temp")

# duplicate a working copy of the cost and revenue layer into scratch
gp.CopyRaster_management(Cost_Layer, "cost")
# get a copy of revenue with value of zero for the regions close to a road
mapAlg = "con(eucdistance(road_rast) < " + str(roadBuffer) + ", 0, " + Revenue_layer + ")"
gp.SingleOutputMapAlgebra_sa(mapAlg, "revenue")

# create layers to track the regions and roads added in each year
gp.SingleOutputMapAlgebra_sa("road_rast - 1", "road_year")
gp.SingleOutputMapAlgebra_sa("setnull(0 == 0, 1)", "region_year")
```

```
##################################################################
## Establish scaling conversion from polyline roads to cells ##
##################################################################

desc = gp.describe(Roads)
roadUnits = desc.spatialreference.linearunitname
del desc

if not rastUnits == roadUnits:
    msg = "Revenue raster and road layers use different units" + gp.GetMessages()
    gp.AddError(msg)
    sys.stderr.write(msg)
    sys.exit(2)

# find the road lengths and density within the road shapefile
# to give an original road density
polylineRoadLength = RoadLength(Roads)

# and the number of road cells that converts to.
roadCells = RoadCellCount('road_rast')

# road length per cell conversion factor
# should be somewhere between cellSize and sqrt(2*cellSize^2)
roadLengthPerCell = polylineRoadLength/roadCells


################################
## Start looping over years ##
################################

for year in (np.arange(numYears)+1):

    ###########################################
    ## Find a target change in road density ##
    ###########################################

    totalRoadLength = RoadCellCount('road_rast') * roadLengthPerCell
    roadDensity = totalRoadLength / Area

    # targetDensityChange is drawn from observed changes from similar original densities
    lowBnd = roadDensity - densityRange
    hiBnd  = roadDensity + densityRange

    # which values in the density array are in bound
    subset = (density > lowBnd) & (density < hiBnd)

    # get the set of matching changes in density, check the set isn't empty
    samples= changedens[subset]
    if len(samples) == 0:
        msg = "No density samples within range of " + str(roadDensity) + gp.GetMessages()
        gp.AddError(msg)
```

```
    sys.stderr.write(msg)
    sys.exit(2)

  # now pick a random density change
  np.random.shuffle(samples)
  targetDensityChange = samples[0]

  # convert that to a road length and get bounds
  newRoadLength = targetDensityChange * Area
  roadUpper = newRoadLength * (1 + roadRange)
  roadLower = newRoadLength * (1 - roadRange)

  # now keep track of road built so far and proposed new road
  # a) built + new < roadLower - build this, build another
  # b) roadLower < built + new < roadUpper - build this one, stop
  # c) roadUpper < built + new - stop immediately
  builtRoad = 0

  # log details for the year
  logHead = (str(year) + '\t' + str(Area) + '\t' + str(totalRoadLength) + '\t' +
             str(targetDensityChange) + '\t' + str(newRoadLength) +'\t')

  # loop until the road building lands within (roadLower, roadUpper)
  while(True):

    # Don't build tiny roads (or zero length roads)
    if (newRoadLength - builtRoad) < roadBuffer:
      log.write(logHead + "\n")
      log.flush()
      break

    ## CHOOSE A ROAD ENDPOINT WITH A GOOD LOCAL RC WITHIN RANGE OF ROADS TO BUILD
    # A) Get the cost distance from the cost layer to the nearest road
    ##   - IDEALLLY use a mask based on euclidean distance from roads to restrict to
    ##     area reachable by remaining road length - more efficient processing
    ##       gp.EucDistance_sa('road_rast', 'costmask', str(remainingRoad))
    ##       gp.Mask = 'costmask' ## no this is not the SA mask. Can't find how to set it
    gp.CostDistance_sa("road_rast", "cost", "cost_dist", "", "cost_bklnk")

    # B) Now get the actual length of the road from each cell to the roads
    #   - convert the distance backlink into a flow backlink
    gp.SingleOutputMapAlgebra_sa("pow(2, cost_bklnk - 1)" , "road_flow_int")
    gp.SingleOutputMapAlgebra_sa("con(road_flow_int < 1 , 0, road_flow_int)" , "road_flow")
    #   - zero values (source - i.e. existing roads) act as sinks and the
    #     rest of the flow link allows distance calculations to those sources
    gp.FlowLength_sa("road_flow", "road_length")

    # D) create an approximate r/c layer for circular concession sized units
    #   - select a concession size for this road
    np.random.shuffle(concessionSizes)
    currConcession = concessionSizes[0]
```

```
        concessionRadius = np.sqrt(currConcession/np.pi)
        concessionStructure = 'Circle ' + str(concessionRadius) + ' MAP'
        # get revenue within each concessionStructure - zero revenue cells deflate local value
        gp.FocalStatistics_sa("revenue",  'patch_rev', concessionStructure, "SUM", "DATA")
        # get extraction cost within each concessionStructure
        gp.FocalStatistics_sa("cost",  'patch_cost', concessionStructure, "SUM", "DATA")
        # get total r/c ratio for each cell including road
        mapAlg = "patch_rev / (patch_cost * " + str(skidderTrack) + " + cost_dist)"
        gp.SingleOutputMapAlgebra_sa(mapAlg, "rcratio")

        # E) Find the area in which the road could fall
        #   - generate a region of cells from roadBuffer out to remaining
        #     road length from existing roads for which revenue > 0
        mapAlg = ("setnull(road_length > " + str(newRoadLength - builtRoad) +
                " | revenue == 0 ,1)")
        gp.SingleOutputMapAlgebra_sa(mapAlg, "road_reg")
        #   - get the eligible area of the rcratio region
        gp.SingleOutputMapAlgebra_sa("rcratio * road_reg", "rcratio_reg")
        gp.Slice_sa("rcratio_reg", "slice_reg", 10, "EQUAL_AREA", 1)
        gp.SingleOutputMapAlgebra_sa("setnull(slice_reg < 10, 1)", "top_reg")

        # F) Get a DBF file of road length and rc ratio in cells within that top region
        gp.RasterToPoint_conversion("top_reg", "target_points.shp", "VALUE")
        gp.sample_sa("road_length; rcratio_reg; cost_dist", "target_points.shp", "region_data.dbf")
        possibleDest = ReadDBF(scratchWorkspace, "region_data.dbf")

        # G) Now choose an end point randomly but weighted by
        #    cumulative sum of rc ratios rescaled into [0,1]
        cumRC = np.cumsum(possibleDest['RCRATIO_RE'])
        cumRC = cumRC / cumRC[-1]
        # random point in [0,1] picks the destination
        RCselect = np.nonzero(cumRC >= np.random.uniform())[0][0]
        proposedRoadLength = possibleDest['ROAD_LENGT'][RCselect]

        ## NOW WE HAVE A DESTINATION - IF THIS DOESN'T GO OVER THE TARGET ROAD LENGTH
        ## THEN RUN THROUGH THE PROCESS OF BUILDING THE ROAD AND MARKING OUT THE
CONCESSION
        if not (builtRoad + proposedRoadLength) > roadUpper:

            # A) Export that point as a pixel
            gp.Select_analysis("target_points.shp", 'target.shp', "POINTID = " + str(RCselect))
            gp.PointToRaster_conversion('target.shp', "POINTID", "road_end", "MOST_FREQUENT",
"NONE", cellSize)

            # B) Get the track of the least cost path to the road network for that region
            gp.CostPath_sa("road_end", "cost_dist", "cost_bklnk", "new_road", "EACH_CELL", "")

            # C) now find a concession around that road head
            #   - find a circular search zone around the road head using
            #     a radius of 2x concession radius, then find most productive unlogged
            #     forest in that area (lower quartile of c/r)
```

```
        #    - note that this can choke if there is too much worked out land
        gp.EucDistance_sa('road_end', 'concZone', str(2 * concessionRadius))
        mapAlg = "setnull(([conczone]  >= 0 & [revenue] > 0)  == 0, cost / revenue)"
        gp.SingleOutputMapAlgebra_sa(mapAlg , 'CellInvRC' )
        gp.CostDistance_sa("road_end","CellInvRC", "rc_dist", "", "")
        gp.Slice_sa("rc_dist", "concSlice", 4, "EQUAL_AREA", 1)
        gp.SingleOutputMapAlgebra_sa("setnull(concSlice > 1, 0)", "concess" )

        # D) get some reporting on the concession
        gp.sample_sa("revenue; cost", "concess", "concess_data.dbf")
        concessionData = ReadDBF(scratchWorkspace, "concess_data.dbf")

        # D) Label new road as 1 and add all the roads together
        gp.SingleOutputMapAlgebra_sa("merge(road_rast, con(new_road >= 0, 1))", "road_final")

        # 12) update the revenue from the concession and close to the road
        gp.SingleOutputMapAlgebra_sa("setnull(road_length > " + str(roadBuffer) + " , 0)",
"workedout")
        gp.SingleOutputMapAlgebra_sa("merge(concess, workedout, revenue)", "rev_final")

        # 13) UPDATE THE REPORTING LAYERS AND LOGFILE
        mapAlg = "merge(road_year, con(new_road >= 0, " + str(year) + "))"
        gp.SingleOutputMapAlgebra_sa(mapAlg, "rd_year_new")
        mapAlg = "merge(region_year, con(concess == 0, " + str(year) + "))"
        gp.SingleOutputMapAlgebra_sa(mapAlg, "reg_year_new")

        if verbose:
           msg = ('Year ' + str(year) + ": building road of " + str(proposedRoadLength) +
                " to concesssion of area " + str(currConcession)+ "\n")
           sys.stdout.write(msg)
           gp.addmessage(msg)

        # - write data on the road to the log file
        concessRoadCost = possibleDest['COST_DIST'][RCselect]
        concessRevenue = np.sum(concessionData['REVENUE'])
        concessExtract = np.sum(concessionData['COST']) * skidderTrack
        concessRC = concessRevenue / (concessRoadCost + concessExtract)

        logEntry = (logHead + str(proposedRoadLength) + '\t' +
                str(concessRoadCost) + '\t' + str(currConcession) + '\t' +
                str(concessRevenue) + '\t' +  str(concessExtract) + '\t' +
                str(concessRC) + "\n")

        log.write(logEntry)
        log.flush()

        # clean up files for next loop
        findBuildFiles = [ "concslice", "cost_bklnk", "cost_dist",
                 "patch_cost", "patch_rev", "rcratio", "rcratio_reg",
                 "region_data.dbf", "region_year", "revenue",
                 "road_flow", "road_flow_int", "road_length",
```

```
                "road_rast", "road_reg", "road_year", "slice_reg",
                "target_points.shp", "top_reg", "conczone",
                "target.shp", "road_end", "new_road", "cellinvrc",
                "rc_dist", "concess", "concess_data.dbf", "workedout"]

        for tmpFile in findBuildFiles:
          gp.Delete_management(tmpFile)

        # rename the remaining files that are needed in the next loop
        gp.rename('rev_final', 'revenue')
        gp.rename('road_final', 'road_rast')
        gp.rename("rd_year_new", "road_year")
        gp.rename("reg_year_new", "region_year")

      else:
        # Tidy up the temporary files from road choice for the next year
        findFiles = [ "concslice", "cost_bklnk", "cost_dist",
                "patch_cost", "patch_rev", "rcratio", "rcratio_reg",
                "region_data.dbf", "road_flow", "road_flow_int", "road_length",
                "road_reg", "slice_reg", "target_points.shp", "top_reg"]

        for tmpFile in findFiles:
          gp.Delete_management(tmpFile)

        break

      # is there enough road unbuilt to go round again
      if (builtRoad + proposedRoadLength) > roadLower:
        break
      else:
        builtRoad += proposedRoadLength

# tidy up
log.close()
```

Appendix D: **The transparency, reliability & utility of land use and land cover change models: an Amazonian case study**

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.

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.*, 2007), 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.*, 2006). 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.

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 (Supporting Information, Appendix 1). 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 S1). 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 S1). 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 S1). 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 (Supporting Information).

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 S1). 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, 2012) 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.

Results

<u>Spatial and temporal scales</u>

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. 1a). 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. 1b), 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). 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. 2). 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).

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).

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.

<u>Drivers of deforestation in the Amazon</u>

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. 3). 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.*, 2006, 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.

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.* (2007) 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.* (2006) 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).

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 (Supporting Information). 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.*, 2007).

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.

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.*, 2007, 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.*, 2001, 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 S1).

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.

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 (Supporting Information), 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, 2012) 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.*, 2007). 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. 4a and Table S2). 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. 4b). A similar pattern was found for Yanai *et al.* (2012) where cumulative model predictions reached 4 – 5% by 2010 (Fig. 4b). 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. 4a and b, respectively) whereas omission errors decreased through time for both cumulative and annual comparisons (Fig. 4a and b). Only one deforestation map was available for Wassenaar *et al.* (2007) 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.* (2007) model tended to have higher omission errors in the eastern Amazon and higher commission errors in the south (Fig. S1), 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. S2 and S3).

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. 4 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.* (2007) (5 km pixel size) model predictions, although for every buffer size these seemed to be consistently lower than Soares-Filho *et al.* (2006) predictions.

Discussion

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.*, 2007, 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).

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 (Supporting Information: 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.

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.

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.

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 (Supporting Information, Appendix 2). 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.

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.

Acknowledgements for Appendix D

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.

References for Appendix D

Aguiar APD, Camara G, Escada MIS (2007) Spatial statistical analysis of land-use determinants in the Brazilian Amazonia: Exploring intra-regional heterogeneity. *Ecological Modelling,* **209**, 169-188.

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

Alcorn JB (1993) Indigenous peoples and conservation. *Conservation Biology,* **7**, 424 - 426.

Araujo C, Bonjean CA, Combes JL, Motel PC, Reis EJ (2009) Property rights and deforestation in the Brazilian Amazon. *Ecological Economics,* **68**, 2461-2468.

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

Betts RA, Malhi Y, Roberts JT (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.

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

Brandão AO, Souza CM (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.

Cantelaube P, Terres JM (2005) Seasonal weather forecasts for crop yield modelling in Europe. *Tellus,* **57**, 476-487.

Carpentier CL, Vosti SA, Witcover J (2000) Intensified production systems on western Brazilian Amazon settlement farms: could they save the forest? *Agriculture Ecosystems & Environment,* **82**, 73-88.

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

Dale VH, Oneill RV, Southworth F, Pedlowski M (1994) Modeling Effects of Land Management in the Brazilian Amazonian Settlement of Rondonia. *Conservation Biology,* **8**, 196-206.

De Koning GHJ, Veldkamp A, Fresco LO (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 GHJ, Verburg PH, Veldkamp A, Fresco LO (1999b) Multi-scale modelling of land use change dynamics in Ecuador. *Agricultural Systems,* **61**, 77-93.

Deadman P, Robinson D, Moran E, Brondizio E (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.

Elith J, H. Graham C, P. Anderson R *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography,* **29**, 129-151.

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

Evans TP, Manire A, De Castro F, Brondizio E, Mccracken S (2001) A dynamic model of household decision-making and parcel level landcover change in the eastern Amazon. *Ecological Modelling,* **143**, 95-113.

Ewers RM, Laurance WF, Souza CM (2008) Temporal fluctuations in Amazonian deforestation rates. *Environmental Conservation,* **35**, 303-310.

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

Ferraz SFD, Vettorazzi CA, Theobald DM, Ballester MVR (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, Jenkins CN, Pimm SL, Keane B, Ross C (2008) Oil and Gas Projects in the Western Amazon: Threats to Wilderness, Biodiversity, and Indigenous Peoples. *PLoS ONE,* **3**, e2932.

Foley JA, Asner GP, Costa MH *et al.* (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 JA, Defries R, Asner GP *et al.* (2005) Global Consequences of Land Use. *Science,* **309**, 570-574.

Forman RTT, Alexander LE (1998) Roads and Their Major Ecological Effects. *Annu. Rev. Ecolo. Syst.,* **29**, 207 - 231.

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

Geist HJ, Lambin EF (2002) Proximate causes and underlying driving forces of tropical deforestation. *Bioscience,* **52**, 143-150.

Gibbs HK, Ruesch AS, Achard F, Clayton MK, Holmgren P, Ramankutty N, Foley JA (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.

Grimm V, Berger U, Bastiansen F *et al.* (2006) A standard protocol for describing individual-based and agent-based models. *Ecological Modelling,* **198**, 115-126.

Huang J, Pontius Jr RG, Li Q, Zhang Y (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.

Inpe (2012) PRODES project - satellite monitoring of the Brazilian Amazon. pp Page.

Killeen TJ, Calderon V, Soria L *et al.* (2007) Thirty Years of Land-cover Change in Bolivia. *Ambio,* **36**, 600 - 606.

Labarta RA, White DS, Swinton SM (2008) Does Charcoal Production Slow Agricultural Expansion into the Peruvian Amazon Rainforest? *World Development,* **36**, 527-540.

Lambin EF (1997) Modelling and monitoring land-cover change processes in tropical regions. *Progress in Physical Geography,* **21**, 375-393.

Lambin EF, Geist HJ, Lepers E (2003) DYNAMICS OF LAND-USE AND LAND-COVER CHANGE IN TROPICAL REGIONS. *Annu. Rev. Environ. Resour.,* **28**, 205 - 241.

Lambin EF, Turner BL, Geist HJ *et al.* (2001) The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change,* **11**, 261-269.

Lapola DM, Schaldach R, Alcamo J, Bondeau A, Koch J, Koelking C, Priess JA (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.

Laurance WF (1999) Reflections on the tropical deforestation crisis. *Biological Conservation,* **91**, 109-117.

Laurance WF, Albernaz AKM, Schroth G, Fearnside PM, Bergen S, Venticinque EM, Da Costa C (2002) Predictors of deforestation in the Brazilian Amazon. *Journal of Biogeography,* **29**, 737-748.

Laurance WF, Cochrane MA, Bergen S *et al.* (2001) The future of the Brazilian Amazon. *Science,* **291**, 438-439.

Le Quere C, Raupach MR, Canadell JG, Marland G, *Et Al.* (2009) Trends in the sources and sinks of carbon dioxide. *Nature Geosci,* **2**, 831-836.

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

Ludeke AK, Maggio RC, Reid LM (1990) An Analysis of Anthropogenic Deforestation Using Logistic-Regression and Gis. *Journal of Environmental Management,* **31**, 247-259.

Maeda EE, De Almeida CM, De Carvalho Ximenes A, Formaggio AR, Shimabukuro YE, Pellikka P (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.

Mann ML, Kaufmann RK, Bauer D *et al.* (2010) The economics of cropland conversion in Amazonia: The importance of agricultural rent. *Ecological Economics,* **69**, 1503-1509.

Mello R, 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.

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

Messina JP, Evans TP, Manson SM, Shortridge AM, Deadman PJ, Verburg PH (2008) Complex systems models and the management of error and uncertainty. *Journal of Land Use Science,* **3**, 11-25.

Messina JP, Walsh SJ (2001) 2.5D Morphogenesis: modeling landuse and landcover dynamics in the Ecuadorian Amazon. *Plant Ecology,* **156**, 75-88.

Millington JA, Perry GW, Romero-Calcerrada R (2007) Regression Techniques for Examining Land Use/Cover Change: A Case Study of a Mediterranean Landscape. *Ecosystems,* **10**, 562-578.

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

Moreira E, Costa S, Aguiar AP, Camara G, Carneiro T (2009) Dynamical coupling of multiscale land change models. *Landscape Ecology,* **24**, 1183-1194.

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

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

Nelson E, Sander H, Hawthorne P *et al.* (2010) Projecting Global Land-Use Change and Its Effect on Ecosystem Service Provision and Biodiversity with Simple Models. *PLoS ONE,* **5**, e14327.

Nepstad D, Soares BS, Merry F *et al.* (2009) The End of Deforestation in the Brazilian Amazon. *Science,* **326**, 1350-1351.

Nepstad DC, Stickler CM, Almeida OT (2006) Globalization of the Amazon soy and beef industries: Opportunities for conservation. *Conservation Biology,* **20**, 1595-1603.

Parker D, Brown D, Polhill JG, Manson SM, Deadman PJ (2008) Illustrating a new 'conceptual design pattern' for agent-based models and land use via five case studies: the MR POTATOHEAD framework. In: *Agent-based Modelleling in Natural Resource Management.* (eds Paredes AL, Iglesias CH) pp Page. Valladolid, Spain, Universidad de Valladolid.

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

Pereira HM, Leadley PW, Proença V *et al.* (2010b) Scenarios for Global Biodiversity in the 21st Century. *Science*.

Pijanowski BC, Brown DG, Shellito BA, Manik GA (2002) Using neural networks and GIS to forecast land use changes: a Land Transformation Model. *Computers, Environment and Urban Systems,* **26**, 553-575.

Polhill JG, Gotts N (2009) Ontologies for transparent integrated human-natural system modelling. *Landscape Ecology,* **24**, 1255-1267.

Pontius Jr RG, Huffaker D, Denman K (2004a) Useful techniques of validation for spatially explicit land-change models. *Ecological Modelling,* **179**, 445-461.

Pontius Jr RG, Shusas E, Mceachern M (2004b) Detecting important categorical land changes while accounting for persistence. *Agric Ecosyst Environ,* **101**, 18-18.

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

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

Rosa IMD, Souza Jr. C, Ewers RM (2012) Changes in size of deforested patches in the Brazilian Amazon *Conservation Biology,* **26**, 932-937.

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

Sarkar S, Crews-Meyer K, Young K, Kelley C, Moffett A (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.

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

Soares-Filho B, Alencar A, Nepstad D *et al.* (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 BS, Cerqueira GC, Pennachin CL (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 BS, Nepstad DC, Curran LM *et al.* (2006) Modelling conservation in the Amazon basin. *Nature,* **440**, 520-523.

Soler LD, Verburg P, Veldkamp A, Escada MIS, Camara G (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.

Steininger MK, Tucker C, Townshend JRG, Killeen TJ, Desch A, Bell V, Ersts P (2001) Tropical deforestation in the Bolivian Amazon *Environ. Conserv.,* **28**, 127–134

Tebaldi C, Knutti R (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 MC, Doblas-Reyes FJ, Mason SJ *et al.* (2006) Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature,* **439**, 576-579.

Verburg PH, Soepboer W, Limpiada R, Espaldon MVO, Sharifa MA, Veldkamp A (2002) Modelling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management,* **30**, 391-405.

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

Versace VL, Ierodiaconou D, Stagnitti F, Hamilton AJ (2008) Appraisal of random and systematic land cover transitions for regional water balance and revegetation strategies. *Agriculture, Ecosystems & Environment,* **123**, 328-336.

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

Walker R, Drzyzga SA, Li YL, Qi JG, Caldas M, Arima E, Vergara D (2004) A behavioral model of landscape change in the Amazon Basin: The colonist case. *Ecological Applications,* **14**, S299-S312.

Walsh SJ, Messina JP, Mena CF, Malanson GP, Page PH (2008) Complexity theory, spatial simulation models, and land use dynamics in the Northern Ecuadorian Amazon. *Geoforum,* **39**, 867-878.

Wassenaar T, Gerber P, Verburg PH, Rosales M, Ibrahim M, Steinfeld H (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.

Wcrp (2012) WCRP CMIP3 Multi-Model Dataset Archive at PCMDI. pp Page.

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

Wu S-S, Qiu X, Usery EL, Wang L (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.

Yanai AM, Fearnside PM, Graça PMLDA, Nogueira EM (2012) Avoided deforestation in Brazilian Amazonia: Simulating the effect of the Juma Sustainable Development Reserve. *Forest Ecology and Management,* **282**, 78-91.

Figure captions

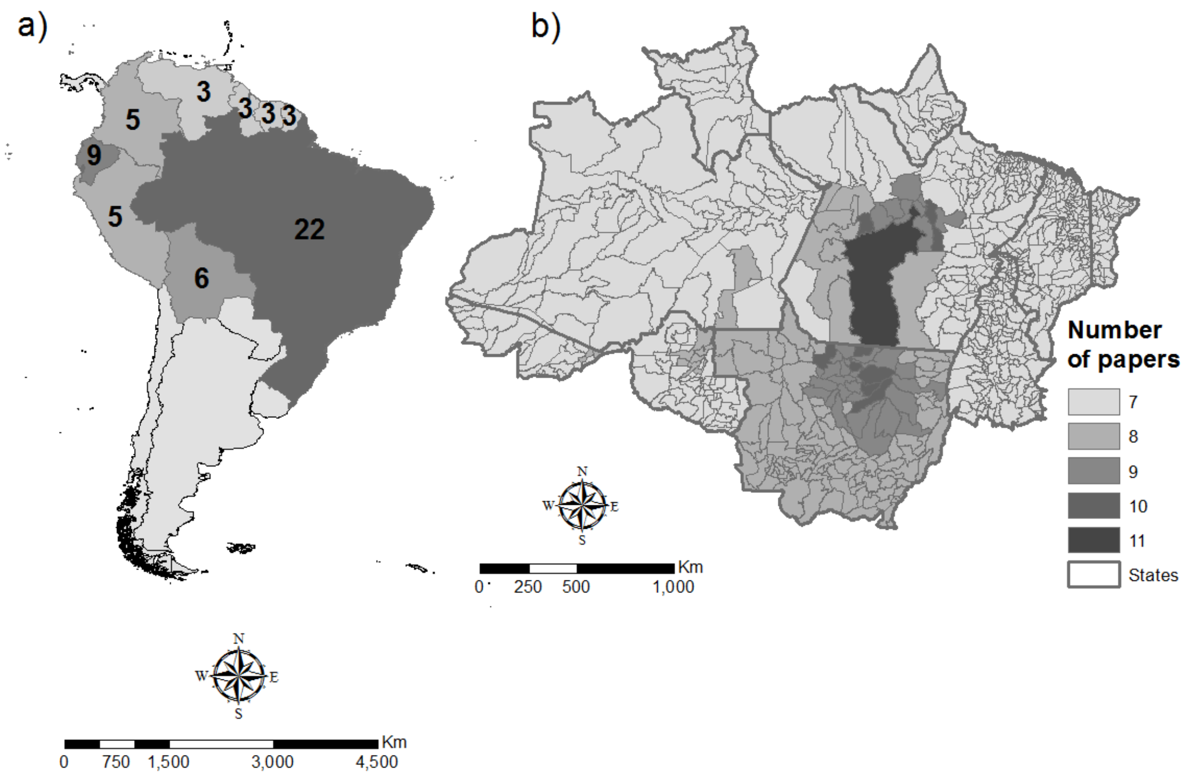


Figure 1 – Number of papers included in our review (a) per country in South America and (b) within the Brazilian Amazon.

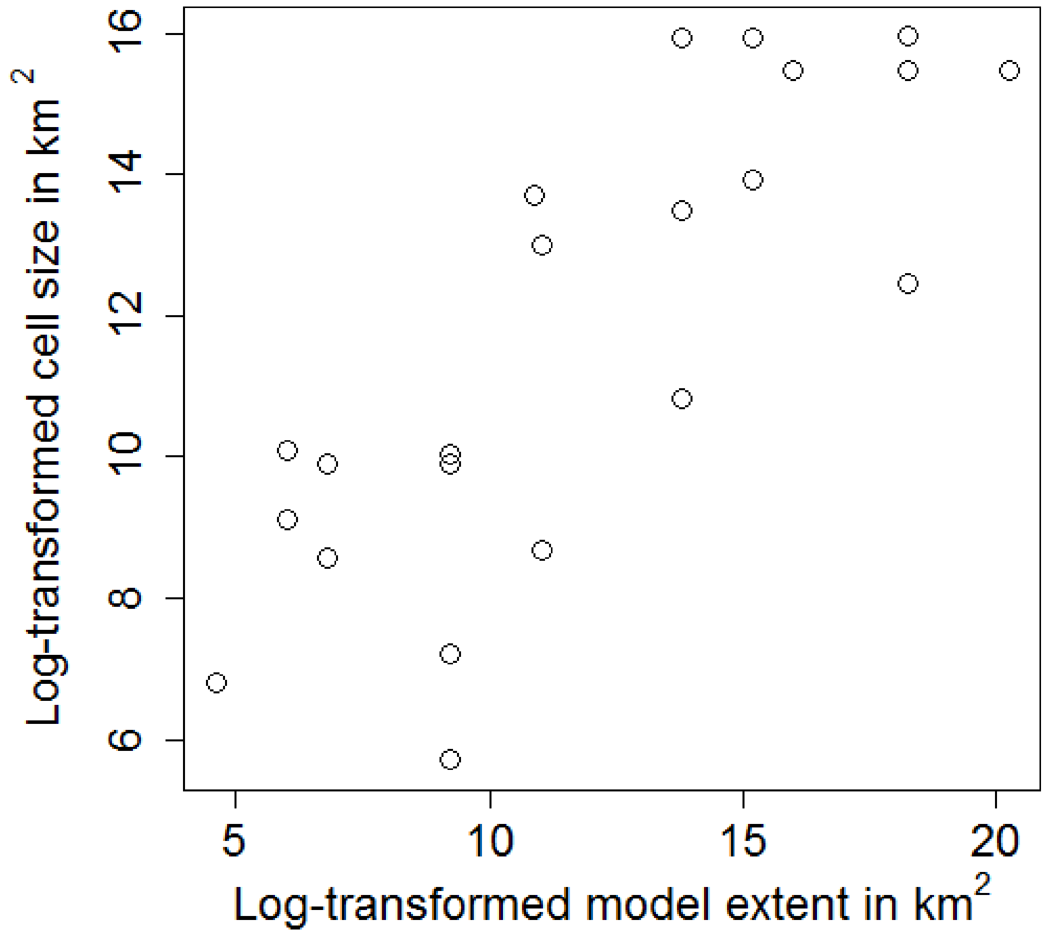


Figure 2 – Correlation between the spatial scale of models (model extent in km^2) and the model resolution (given by the cell/pixel area in km^2).

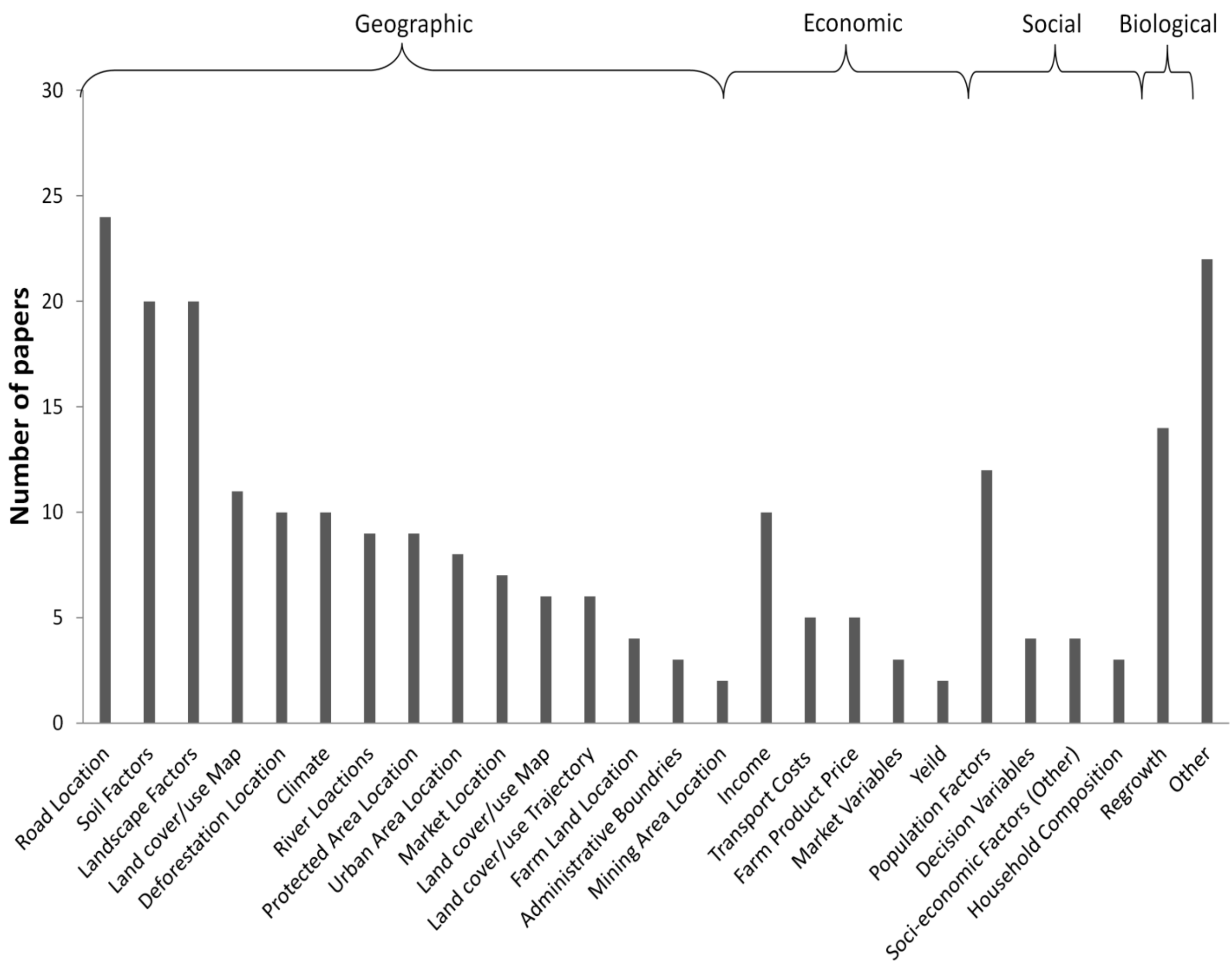


Figure 3 – The number of papers using each input type. Inputs are divided according to class; geographic, economic, social or biological. Soil factors have been put into a single group that consists of factors such as soil moisture and soil texture. Landscape factors include inputs such as altitude and 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.

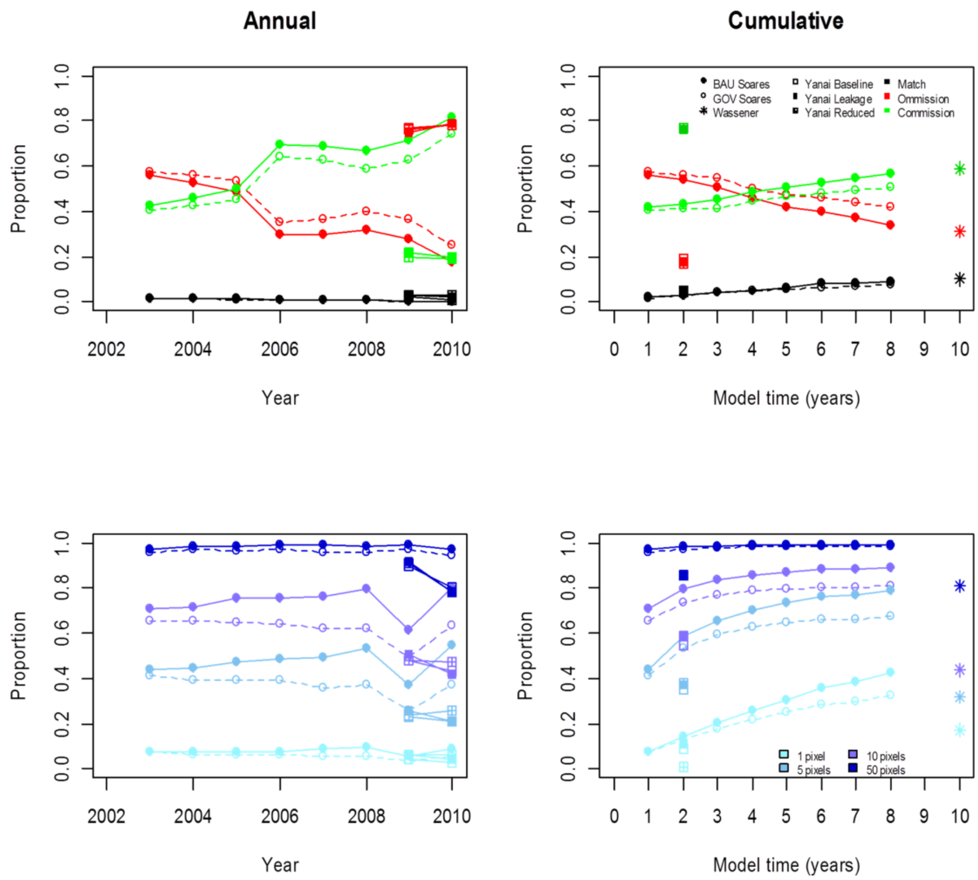


Figure 4 – Pixel-by-pixel comparisons between observed deforestation and predictions made (a) annually and (b) cumulatively by Soares-Filho *et al.* (2006) at 1 x 1 km grid cells, from both governance (GOV) and business-as-usual (BAU) scenarios between 2003 and 2010, and by Yanai *et al.* (2012) for 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).

Supporting Information

Methodological results

The methodological details extracted from the 35 papers shown in Appendix 1 are shown in Table S1. 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.

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.*, 2007, 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.

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.

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.*, 2008a), 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.*, 2008a). This model attempts to recreate the road building decisions made by the logging industry (Arima *et al.*, 2008a), 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.*, 2005, Arima *et al.*, 2008b). In the original model, Arima *et al.* (2005) 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.* (2008b) model only destination determinate roads were considered.

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.*, 2007, 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.*, 2007, 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.* (2001) 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.

The difficulties of model validation

The most frequently used method of model validation is the Receiver Operating Characteristic (ROC) curve (Wassenaar *et al.*, 2007, 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).

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.*, 2007, 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 S2. 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, figures S1 shows the comparisons made between PRODES observed deforestation between 2000 and 2010 and predicted made by Wassenaar *et al.* (2007); figure S2 and S3 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.*, 2007), because Lapola *et al.* (2010) and Maeda *et al.* (2011) only provided outputs for 2020 and 2015, respectively.

References for Appendix D Supporting material

Arima EY, Walker RT, Perz SG, Caldas M (2005) Loggers and Forest Fragmentation: Behavioral Models of Road Building in the Amazon Basin. *Annals of the Association of American Geographers,* **95**, 525-541.

Arima EY, Walker RT, Sales M, Souza Jr. C, Perz SG (2008a) The Fragmentation of Space in the Amazon Basin: Emergent Road Networks. *Photogrammetric Engineering & Remote Sensing,* **74**, 699 - 709.

Arima EY, Walker RT, Sales M, Souza Jr. C, Perz SG (2008b) The Fragmentation of Space in the Amazon Basin: Emergent Road Networks. *Photogrammetric Engineering & Remote Sensing,* **74**, 699–709.

Barlow J, Ewers RM, Anderson L *et al.* (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.

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

Dale VH, Oneill RV, Southworth F, Pedlowski M (1994) Modeling Effects of Land Management in the Brazilian Amazonian Settlement of Rondonia. *Conservation Biology,* **8**, 196-206.

De Koning GHJ, Veldkamp A, Fresco LO (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 GHJ, Verburg PH, Veldkamp A, Fresco LO (1999b) Multi-scale modelling of land use change dynamics in Ecuador. *Agricultural Systems,* **61**, 77-93.

Deadman P, Robinson D, Moran E, Brondizio E (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.

Eastman JR, Van Fossen ME, Solarzano LA (2005) Transition potential modeling for land cover change. . In: *GIS, Spatial Analysis and Modeling.* (eds Maguire D, Batty M, Goodchild M) pp Page. Redlands, CA, ESRI Press.

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

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

Fearnside PM (2008) The Roles and Movements of Actors in the Deforestation of Brazilian Amazonia. *Ecology and Society,* **13**, -.

Ferraz SFD, Vettorazzi CA, Theobald DM, Ballester MVR (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.

Frohn RC, Mcgwire KC, Dale VH, Estes JE (1996) Using satellite remote sensing analysis to evaluate a socio-economic and ecological model of deforestation in Rond√¥nia, Brazil. *International Journal of Remote Sensing,* **17**, 3233-3255.

Geist HJ, Lambin EF (2002) Proximate causes and underlying driving forces of tropical deforestation. *Bioscience,* **52**, 143-150.

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*.

Lapola DM, Schaldach R, Alcamo J, Bondeau A, Koch J, Koelking C, Priess JA (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 DM, Schaldach R, Alcamo J *et al.* (2011) Impacts of Climate Change and the End of Deforestation on Land Use in the Brazilian Legal Amazon. *Earth Interactions,* **15**, 1-29.

Laurance WF, Cochrane MA, Bergen S *et al.* (2001) The future of the Brazilian Amazon. *Science,* **291**, 438-439.

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

Maeda EE, De Almeida CM, De Carvalho Ximenes A, Formaggio AR, Shimabukuro YE, Pellikka P (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.

Mann ML, Kaufmann RK, Bauer D *et al.* (2010) The economics of cropland conversion in Amazonia: The importance of agricultural rent. *Ecological Economics,* **69**, 1503-1509.

Mena CF, Walsh SJ, Frizzelle BG, Xiaozheng Y, Malanson GP (2011) Land use change on household farms in the Ecuadorian Amazon: Design and implementation of an agent-based model. *Applied Geography,* **31**, 210-222.

Messina JP, Walsh SJ (2001) 2.5D Morphogenesis: modeling landuse and landcover dynamics in the Ecuadorian Amazon. *Plant Ecology,* **156**, 75-88.

Michalski F, Peres CA, Lake IR (2008) Deforestation dynamics in a fragmented region of southern Amazonia: evaluation and future scenarios. *Environmental Conservation,* **35**, 93-103.

Moreira E, Costa S, Aguiar AP, Camara G, Carneiro T (2009) Dynamical coupling of multiscale land change models. *Landscape Ecology,* **24**, 1183-1194.

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

Nepstad D, Soares BS, Merry F *et al.* (2009) The End of Deforestation in the Brazilian Amazon. *Science,* **326**, 1350-1351.

Pfaff A, Robalino J, Walker R *et al.* (2007) Roads and deforestation in the Brazilian Amazon. *Journal of Regional Science,* **47**, 109-123.

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

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

Soares-Filho B, Alencar A, Nepstad D *et al.* (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 BS, Cerqueira GC, Pennachin CL (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 BS, Nepstad DC, Curran LM *et al.* (2006) Modelling conservation in the Amazon basin. *Nature,* **440**, 520-523.

Soler LD, Verburg P, Veldkamp A, Escada MIS, Camara G (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.

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.

Verburg PH, Soepboer W, Limpiada R, Espaldon MVO, Sharifa MA, Veldkamp A (2002) Modelling the spatial dynamics of regional land use: The CLUE-S model. *Environmental Management,* **30**, 391-405.

Walker R, Drzyzga SA, Li YL, Qi JG, Caldas M, Arima E, Vergara D (2004) A behavioral model of landscape change in the Amazon Basin: The colonist case. *Ecological Applications,* **14**, S299-S312.

Walsh SJ, Messina JP, Mena CF, Malanson GP, Page PH (2008) Complexity theory, spatial simulation models, and land use dynamics in the Northern Ecuadorian Amazon. *Geoforum,* **39**, 867-878.

Wassenaar T, Gerber P, Verburg PH, Rosales M, Ibrahim M, Steinfeld H (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.

Yanai AM, Fearnside PM, Graça PMLDA, Nogueira EM (2012) Avoided deforestation in Brazilian Amazonia: Simulating the effect of the Juma Sustainable Development Reserve. *Forest Ecology and Management,* **282**, 78-91.

Figure captions

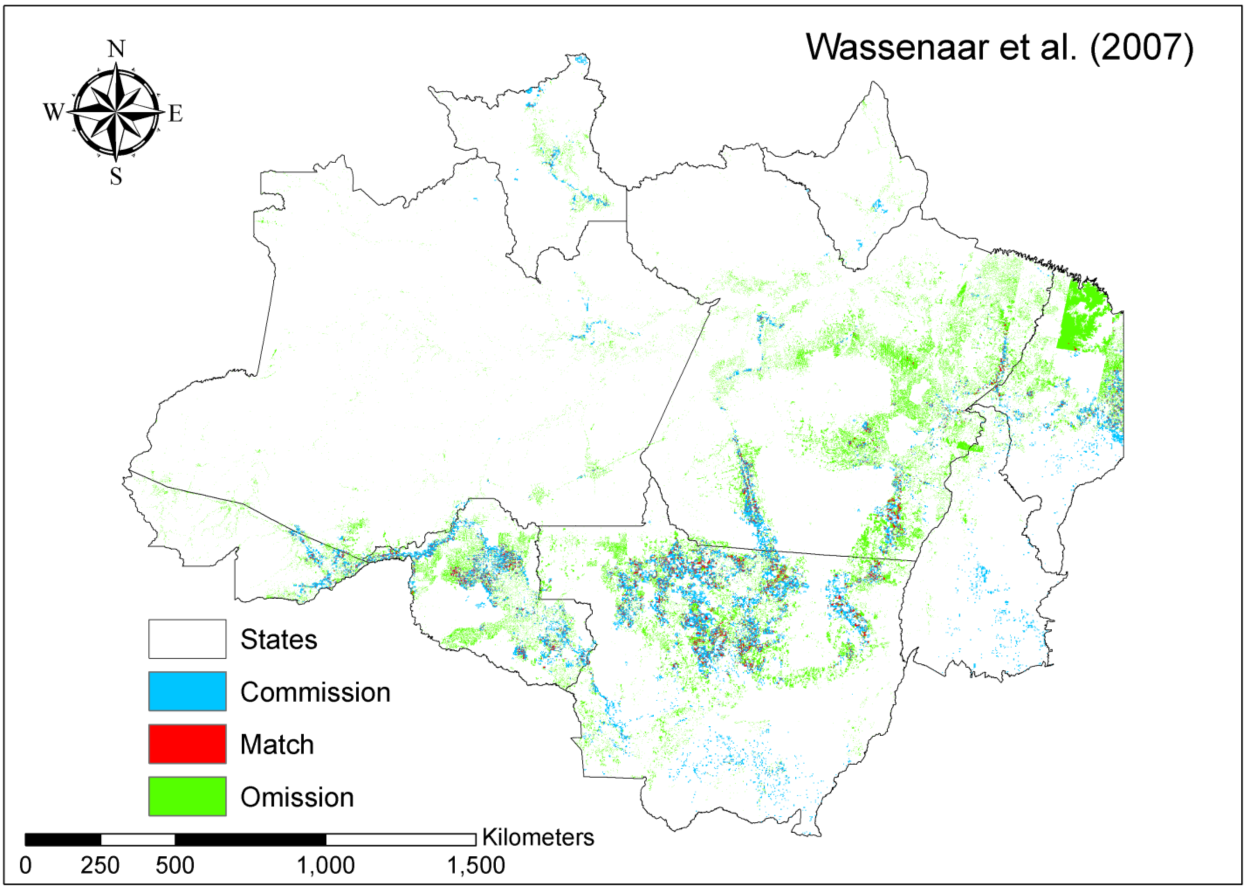


Figure S1 – Match, omission and commission between observed data from PRODES and model predictions made by Wassenaar *et al.* (2007) from 2000 to 2010.

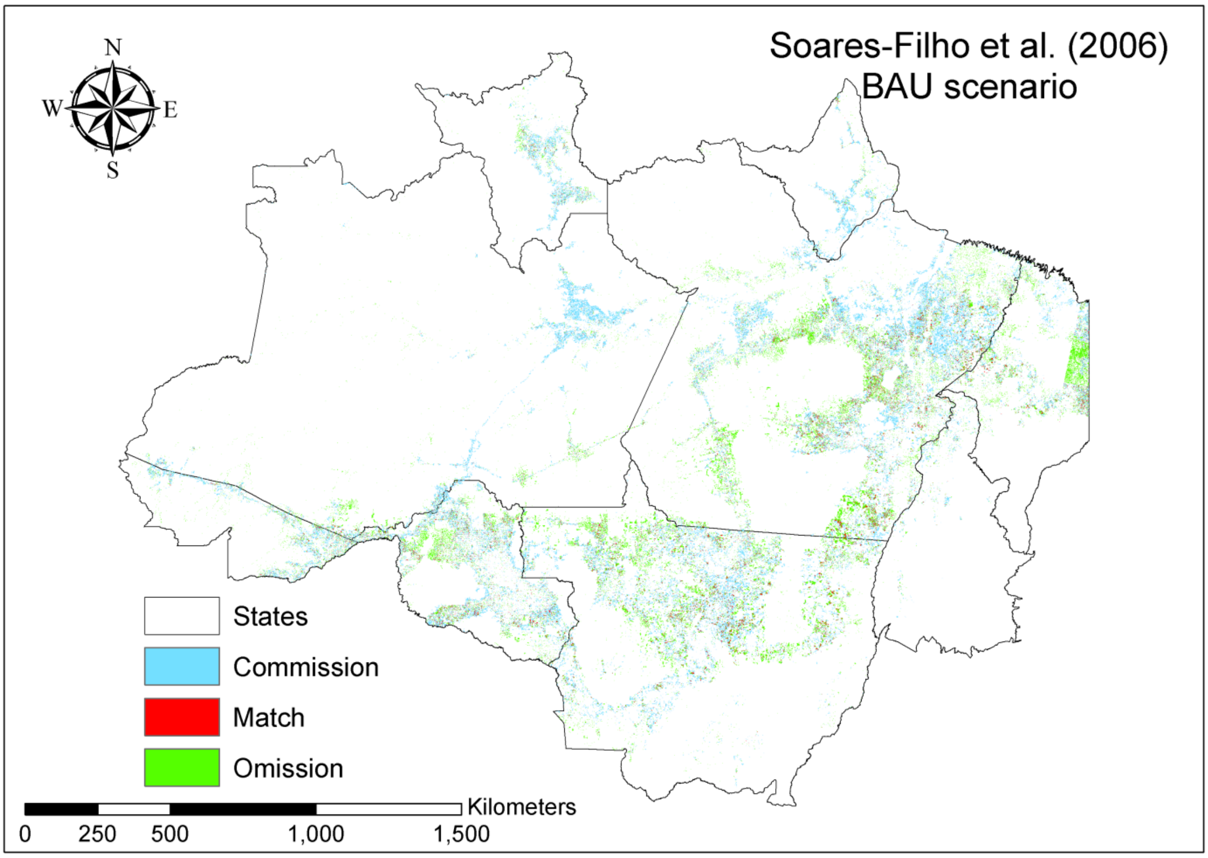


Figure S2 – Match, omission and commission between observed data from PRODES and model predictions by Soares-Filho *et al.* (2006) for the business-as-usual (BAU) scenario from 2002 to 2010.

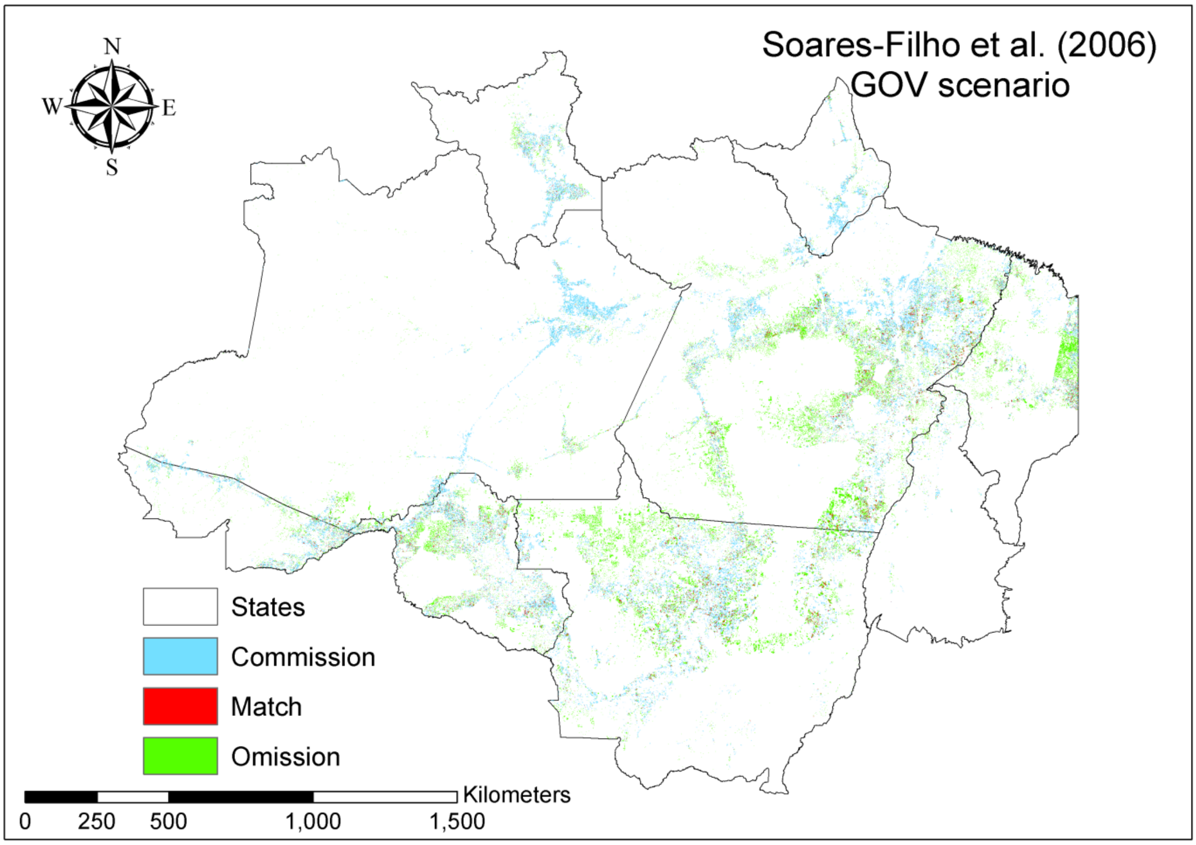


Figure S3 – Match, omission and commission between observed data PRODES and model predictions by Soares-Filho *et al.* (2006) for the governance (GOV) scenario from 2002 to 2010.

Tables

Table S1 – 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	Scenarios	Region	If Brazil, state	Cell size*
Carpentier *et al.* (2000)	FaleBem	Optimisation Agent-based	farm	dynamic	deterministic	4	Brazil	Acre	not spatial
Dale *et al.* (1994)	DELTA	Agent-based	farm & individual	dynamic	stochastic	3	Brazil	Rondônia	100 m
de Barros *et al.* (2005)	n/a	Other (Markov chain)	n/a	dynamic	deterministic	3	Brazil	Rondônia	unclear
de Koning *et al.* (1999b)	CLUE	Statistic	n/a	dynamic	deterministic	1	Ecuador	n/a	5 min
de Koninget al. (1999a)	CLUE	Statistic	n/a	dynamic	deterministic	6	Ecuador	n/a	5 min

de Souza Soler *et al.* (2007)	CLUE-S	Statistic	n/a	static	stochastic	1	Brazil	Rondônia	250 m
Deadman *et al.* (2004)	LUCITA	Agent-based	farm	static	stochastic	1	Brazil	Pará	1 ha
Etter *et al.* (2006b)	IDRISI	Statistic	n/a	static	deterministic	1	Colombia	n/a	1 ha
Etter *et al.* (2006a)	n/a	Statistic	n/a	static	stochastic	1	Colombia	n/a	2 km
Evans *et al.* (2001)	n/a	Agent-based	parcel	dynamic	stochastic	2	Brazil	Pará	unclear
Labarta *et al.* (2008)	n/a	Optimisation Agent-based	household	dynamic	deterministic	6	Peru	n/a	not spatial
Lapola *et al.* (2011)	LandSHIFT	Other (hierarchal)	n/a	dynamic	deterministic	4	Brazil	All	5 arc
Lapola *et al.* (2010)	LandSHIFT	Other (hierarchal)	n/a	dynamic	deterministic	4	Brazil	All	5 arc
Laurance *et al.* (2001)	n/a	Statistic	n/a	static	deterministic	2	Brazil	All	unclear
Lopez *et al.* (2010)	n/a	Agent-based	household	static	deterministic	1	Ecuador	n/a	20 m
Maeda *et al.* (2010)	DINAMICA	CA	n/a	dynamic	stochastic	2	Brazil	Mato Grosso	100 m
Mann *et al.* (2010)	n/a	Statistic	n/a	static	deterministic	1	Brazil	Mato Grosso	232 m
Mena *et al.* (2011)	n/a	Agent-based	farm	dynamic	stochastic	1	Ecuador	n/a	unclear
Messina & Walsh (2001)	n/a	CA	farm	dynamic	stochastic	1	Ecuador	n/a	30 m
Michalski *et al.* (2008)	IDRISI	Other (Neural Network)	n/a	dynamic	deterministic	1	Brazil	Mato Grosso	unclear
Moreira *et al.*	TerraME	Other	farm	dynamic	deterministic	2	Brazil	ALL	5 km

(2009)		(hierarchical) Agent-based						Pará	1 km
Muller *et al.* (2011)	n/a	Statistic	n/a	dynamic	stochastic	1	Bolivia	n/a	200m
Nepstad *et al.* (2010)	DINAMICA	CA	n/a	dynamic	stochastic	3	Brazil	All	unclear
Sarkar *et al.* (2009)	n/a	CA Agent-based	n/a	dynamic	stochastic	1	Peru	n/a	30 m
Soares-Filho *et al.* (2002)	DINAMICA	CA	n/a	dynamic	stochastic	1	Brazil	Mato Grosso	100 m
Soares-Filho *et al.* (2004)	DINAMICA	CA	n/a	dynamic	stochastic	4	Brazil	N. Mato Grosso S. Pará	250 m
Soares-Filho *et al.* (2006)	DINAMICA	CA	n/a	dynamic	stochastic	8	Brazil	All	1 km
Vadez *et al.* (2008)	n/a	Agent-based	household	static	deterministic	5	Bolivia	n/a	not spatial
Walker *et al.* (2004)	n/a	Agent-based	farm	dynamic	stochastic	1	Brazil	Pará	20 m
Walsh *et al.* (2008)	n/a	CA Agent-based	farm	dynamic	stochastic	2	Ecuador	n/a	30 m
Wassenaar *et al.* (2007)	CLUE	Statistic	n/a	dynamic	deterministic	1	South America	n/a	5 km
Silvestrini *et al.* (2011)	DINAMICA	CA	n/a	dynamic	stochastic	3	All Amazon	n/a	2 km
Sangermano *et al.* (2012)	IDRISI	Other (Neural Network)	n/a	dynamic	deterministic	3	Bolivia	n/a	1 km
Mello & Hildebrand (2012)	n/a	Optimisation	farm	static	deterministic	3	Brazil	Pará	not spatial
Yanai *et al.* (2012)	DINAMICA	CA	n/a	dynamic	stochastic	3	Brazil	Amazonas	250 m

9

Continuation of Table S1

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, distance to

(2011)	economic	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.* (2004)	bio geo	landscape map, vegetation, soil, altitude, slope, protected areas, distance to main and 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

13 **Continuation of Table S1**

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.* (2011)	maps and trajectory	analytic hierarchy process / multi-criteria suitability analysis	fuzzy similarity index kappa classification	1	44
Lapola *et al.*	land-use map and	analytic hierarchy process /	ROC	1	17

(2010)	livestock density maps	multi-criteria suitability analysis			
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

Table S2 – 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**
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

(continuation of Table S2)

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

Appendix 1 For Appendix D

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.

Appendix 2 For Appendix D

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.* (2007) 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).