MODELLING THE SEVERITY OF DEFORESTATION IN COASTAL TANZANIA

AND A

COMPARISON OF DATA SOURCES

Oliver Griffin

September 2012

00713294

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science and the Diploma of Imperial College London

DECLARATION OF OWN WORK

I declare that this thesis

"Modelling the Severity of Deforestation in Coastal Tanzania and a Comparison of Data Sources"

is entirely my own work and that where material could be construed as the work of others, it is fully cited and referenced, and/or with appropriate acknowledgement given.

Signature:

Name of student: Oliver Griffin

Name of supervisors: Dr. Neil Burgess

Dr. Lauren Coad

Dr. David Orme

Dr. Peter Long

TABLE OF CONTENTS

LIST OF ACRONYMS

AIC	Akaike Information Criterion
CBD	Convention on Biological Diversity
CI	Conservation International
EOS	Earth Observing System
ETM+	Enhanced Thematic Mapper Plus
GAM	Generalized Additive Models
GDP	Gross Domestic Product
GLM	Generalised Linear Model
IGBP	International Geosphere-Biosphere Programme
IUCN	International Union for Conservation of Nature
MLC	Maximum Likelihood Classification
MOD12Q1	MODIS global land cover product (500 m resolution)
MOD44B	MODIS Vegetation Continuous Fields product (250 m resolution)
MODIS	Moderate-resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration
NBS	National Bureau of Statistics
NEMC	National Environmental Management Council
REDD+	Reducing Emissions from Deforestation and Forest Degradation in Developing Countries
SUA	Sokoine University of Agriculture
UNFCCC	United Nations Framework Convention on Climate Change
UTM	Universal Transverse Mercator (coordinate system)
VCF	Vegetation Continuous Fields
WDPA	World Database on Protected Areas
WWF	World Wide Fund for Nature

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The coastal forests of Tanzania have typically been overlooked in favour of the more spectacular Eastern Arc Mountains. Closer examination has revealed a unique and diverse ecosystem, home to exceptional levels of endemism across many major taxa, distributed heterogeneously across several hundred forest patches. These forest remnants are threatened by further fragmentation, degradation, and deforestation.

Model development can be used to elucidate the drivers of deforestation, predict the location of future deforestation, produce scenarios of future deforestation rates, inform the design of government policy, and provide a baseline against which to test for additionality in programmes such as REDD+. It can also be used to analyse protected area effectiveness.

Here I examine 3 sources of remotely sensed data (supplemented Landsat data, MOD12Q1 data, and MOD44B data) and their suitability for use in deforestation models. To determine drivers of deforestation severity in coastal Tanzania, I use generalised additive models (GAMs) to describe the non-linear relationship between deforestation severity and a set of geographic, biogeographic, and socioeconomic data.

I find that drivers differ across varying initial forest cover, with this factor being the largest predictor of deforestation severity. The explanatory power of the models produced was low, and I discuss reasons for this – including the use of a novel response variable.

Word count: 11,912

ACKNOWLEDGEMENTS

Firstly, thank you to my supervisors for overseeing this thesis: to Dr. Lauren Coad for helping me conceptualise the project and for her encouragement; to Dr. Peter Long for his assistance with remotely sensed data and his hospitality in Oxford; to Dr. David Orme for his patience whilst teaching me spatial modelling and for making horribly complex things seem simple (for a few fleeting moments, at least); and to Dr. Neil Burgess for his invaluable knowledge of the realities of Tanzania and coastal forests. Thanks also to Karyn Tabor of Conservation International and Japhet Kashaigili of the Sokoine University of Agriculture for their data and advice, and to Jonathon Green for his modelling guidance and suggestions.

Thank you to all those at Silwood Park: the students for their company, and those in charge for making Silwood such an enjoyable place to be.

Thank you, as always, to my parents for their unyielding support.

Harri: *misaotra betsaka takolaka mamy*.

And thank you to Uncle Fred and Aunty Fan. I can never thank you in person, but without your kindness my education would not have been possible.

1 INTRODUCTION

Despite tropical forests supporting more than 50% of all described species (Dirzo & Raven 2003) and playing a disproportionately important role in global carbon storage and energy cycles (IPCC 2007), the rate at which they are destroyed continues to be dismayingly high. In each year of the last decade, an average of 9.34 million hectares of forest were lost from the tropics, down from 11.33 million hectares each year in the 1990s (FRA 2010). While the rate of net loss of forest appears to be falling in some areas (FRA 2010), this is due in part to afforestation efforts, especially in China and Europe, which can produce monoculture plantations not comparable with the richness of natural forest ecosystems - indeed, they have been described as 'green deserts' (Acosta 2011). A proportion of the falling deforestation rates in developed countries can be explained by the 'outsourcing' of logging and timber extraction to developing countries, through the importation of both legal and illegal timber products (Meyfroidt et al. 2010).

1.1 NEGATIVE EFFECTS OF DEFORESTATION

1.1.a Species loss

It has been suggested that we are in the midst of a sixth great extinction, with species extinction rates currently 100 to 1000 times higher than baseline rates (Barnosky et al. 2011). Habitat loss is the leading cause of species extinction (Pimm & Raven 2000) and, with the removal of their habitat through deforestation, forest-dependent species stand little chance of survival. More than 20% of all terrestrial species assessed on the International Union for Conservation of Nature's (IUCN) Red List are threatened by logging & wood harvesting (IUCN 2011), while numerous other factors contribute to deforestation, such as clearing forest and woodland for agricultural, infrastructural, and settlement expansion (Rademaekers et al. 2010).

1.1.b Climate change

Regional changes induced by deforestation can be dramatic, including meteorological, hydrological, and biological alterations. Models for the Amazon basin suggest that, following deforestation, significant increases in mean surface temperature and decreases in precipitation, evapotranspiration, and runoff are likely (Nobre et al. 1991). Increases in the length of the dry season are also expected, which can hinder the reestablishment of tropical forest and lead to encroachment of savannah into previously forested regions (Nobre et al. 1991). Similar effects were predicted in models produced by Eltahir & Bras (1994), who also note increased albedo as a result of deforestation. Bradshaw et al. (2007) find deforestation increases flood risk and severity in the developing world, and that reforestation may reduce the frequency of floods and the severity of flood-related disasters.

Globally, forests store 289 gigatonnes (Gt) of carbon in their biomass alone (FRA 2010), comparable with the 337 Gt carbon released to the atmosphere from the burning of fossil fuels and cement production globally since 1751 (Boden et al. 2010). This stored total is being reduced by an estimated 0.5 Gt per year (FRA 2010). Deforestation currently contributes around 12% of total anthropogenic carbon dioxide emissions (Van der Werf et al. 2009), and is second only to fossil fuels. Deforestation and degradation contribute to emissions through the combustion of forest biomass, the decomposition of plant material dependent on forest cover, and the release of carbon stored in eroded soils previously held in place by forests (Van der Werf et al. 2009).

Given the significant contribution to anthropogenic greenhouse gas emissions that deforestation makes, mechanisms that aim to reduce these emissions are being considered (Gullison et al. 2007). One such programme is Reducing Emissions from Deforestation and Forest Degradation in Developing Countries (REDD+), which includes biological conservation, sustainable management of forests, and augmentation of forest carbon stores. Based principally on the valuation and trade of forest carbon, this controversial framework fits well within the pervasive neo-liberal (Castree 2010) market-based approach to environmental conservation being promoted globally by developed countries (e.g. the United States of America (Jenkins 2007)).

The REDD+ mechanism is a comprehensive agreement under the UN Framework Convention on Climate Change (UNFCCC) and is hoped to become operational within the next few years in some areas (Burgess et al. 2010), though the timeframe varies from country to country (Wertz-Kanounnikoff & Angelsen 2009).

1.2 MODELLING DEFORESTATION

The impetuses behind developing models of deforestation are numerous. Model development can be used to elucidate the drivers of deforestation (Dávalos et al. 2011), predict the location of future deforestation (Schneider & Pontius Jr. 2001), and to produce scenarios of future deforestation rates (Soares-Filho et al. 2006). It can be used to inform the design of government policy (Lambin 1994) and to provide a baseline against which the effect of policies can be measured - vital in the assessment of additionality (or 'avoided deforestation') for programmes such as REDD+, supporting the mitigation of climate change through reduced deforestation (Santilli et al. 2005; Scharlemann et al. 2010). It can also be used to analyse protected area effectiveness – essential as protected areas form a cornerstone of conservation (Adams 2004). The Convention of Biological Diversity's (CBD) Aichi Biodiversity Target 11 sets a goal of 17% terrestrial and inland water areas, and 10% of coastal and marine areas to be conserved effectively by 2020 (CBD 2010). Target 5 aims for a halving of habitat loss, including forests, by 2020. Without measures of protected area effectiveness or habitat extent, no meaningful assessment of whether these targets have been met can be made (Chape et al. 2005).

Remotely sensed satellite data is the most commonly used technique for measuring deforestation and, in some cases (due to inaccessibility and impracticability of aerial surveys), the only practical approach (Tucker & Townshend 2000). A wide spectrum of remotely sensed data products exist, with varying resolution, availability, and cost. Selection of an appropriate data product is essential to produce valid, cost-effective conclusions. Where medium resolution data (250 m - 1 km) may be appropriate for annual monitoring of large events, the detection of fragmentation and the removal of small forest patches may require higher resolution data (10 m - 60 m) (Horning et al. 2010). The conservation field is limited in funds, and appropriate allocation of resources is important (Nichols & Williams 2006). The cost of remotely sensed data spanning a country can range from free to >£10,000 (before ground-truthing and validation), and so selection of cost-effective data is important (Achard 2006).

1.3 AIMS AND OBJECTIVES

The objective of this project was to construct a model of deforestation severity in coastal Tanzania, with the aim of determining drivers of deforestation and quantifying their relative contribution to forest loss. This will support larger Tanzanian and coastal forest conservation efforts, highlighting threats to be addressed with conservation action. It will also lay the groundwork for a future assessment of protected area effectiveness in the coastal region, improving the understanding of management inputs and of what makes some protected areas more effective than others (Joppa & Pfaff 2010). I will compare sources of land cover change data to assess the necessity of using labour intensive and costly data products over freely available but coarser resolution products. I will also examine the extent of deforestation in the region, and make a preliminary assessment of deforestation rates inside and outside protected areas. I aim to:

- Quantify rates of forest and woodland loss in the coastal region of Tanzania using supplemented Landsat data.
- Compare supplemented Landsat data with freely available Moderate-resolution Imaging Spectroradiometer (MODIS) products: MOD12Q1 (a global land cover product with 500 m resolution), and MOD44B (Vegetation Continuous Fields product with a 250 m resolution).
- Use supplemented Landsat derived variables along with geographic, biogeographic, and socioeconomic data to produce a generalised additive model (GAM) of deforestation severity in coastal Tanzania.

2 BACKGROUND

2.1 STUDY LOCATION

2.1.a Tanzania

Tanzania, on Africa's eastern coast, is home to ~42,750,000 people, with an economy based on gold production, tourism, and agriculture, the latter of which alone accounts for around one third of Tanzania's GDP, 85% of its exports, and employs around 80% of the work force (CIA 2012). The country's political boundaries encompass immense biological richness, most famously the Eastern Arc Mountains. Tanzania hosts portions of six of the top 34 global biodiversity hotspots, and both the Eastern Arc and coastal forests of Tanzania are part of one hotspot defined by Mittermeier et al. (2005). Together these two regions contain the smallest remaining area of habitat of any of these hotspots (Brooks et al. 2002). They are also part of one of the 200 Global Ecoregions defined by WWF (Olson et al. 2001) - the Northern Zanzibar-Inhambane Coastal Forest Mosaic ecoregion. The coastal region is a Class I conservation priority under the biodiversity conservation prioritisation scheme for Africa developed by Burgess et al. (2006), as an ecoregion with globally important biological value facing high threats.

A total of 38% (~335,000 km^2) of terrestrial Tanzania is covered by forest, with 78% of that total area designated for timber production, 24% for multiple uses, and 6% for biodiversity conservation (FRA 2010). Total forest area was destroyed at a rate of 1.16% per year in the period 2005-2010, and during this time Tanzania had the fifth highest annual net loss of forest area in the world, losing 4,030 km^2 each year (FRA 2010).

Across the whole of Tanzania, natural systems receive protection from more than 600 designated protected areas, with management responsibilities ranging from village-level to central government, and with goals ranging from provision of sustainable firewood for local communities to access-controlled nature reserves (IUCN & UNEP 2012). Tanzania, along with Zambia, has the highest proportion of its land designated as protected in Africa - around 40% (Veit & Benson 2004). The establishment of protected areas first began during the German occupation of East Africa (1884-1919), during which time colonial laws protecting game and forest were brought into force, at the expense of traditional practices of hunting, firewood collection and cattle grazing (Goldstein 2004). The WDPA reports 582 terrestrial and 26 marine protected areas across Tanzania (IUCN & UNEP 2012).

Recent policy approaches to conservation include the establishment of the National Environmental Management Council (NEMC) and the passing of the Environmental Management Act 20, of 2004. Pallangyo (2007) provides an extensive review of environmental law in Tanzania. This country is a participant in the World Bank's Forest Carbon Partnership Facility, is one of 9 pilot countries for the UN-REDD mechanism, and receives significant funding from the Norwegian, German, and Finnish governments (Burgess et al. 2010).

2.1.b **Coastal forest**

Precise definition of the coastal forests of eastern Africa is complex, with a wide range of views of their geographical extent, biological affinities, and their place in wider vegetation formation types (Burgess & Clarke 2000). The coastal forests of eastern Africa are broadly synonymous with the forests of the Zanzibar-Inhambane regional mosaic (White 1983). Burgess & Clarke (2000) provide a formal definition which includes coastal dry forest, coastal scrub forest, coastal *Brachystegia* forest, coastal riverine/groundwater/swamp forest, and coastal/Afromontane transition forest. Degraded scrub and thicket will regenerate into forest, and so from a conservation perspective the coastal forests need not be defined too rigidly (Burgess & Clarke 2000).

The coastal forests of Eastern Africa were once poorly studied and overshadowed by the more accessible and spectacular Eastern Arc Mountains (Burgess & Clarke 2000). Closer examination has revealed a unique and diverse ecosystem, home to exceptional levels of endemism across many major taxa.

These coastal forests are composed of more than 350 forest patches (Burgess & Clarke 2000; Burgess et al. 2003), with most differing in their community structure and species composition. This forest mosaic covers primarily Mozambique (4,778 km^2), Kenya (787 km^2), and Tanzania (692 km^2)(Burgess et al. 2003). The total area of the coastal forests (6,259 km^2) supports 554 endemic species of plant and 52 endemic vertebrate species – giving an average of 9.6 endemic plant or vertebrate species per 100 km^2 of habitat. Comparing this with the corresponding value of the non-forested vegetation of the coastal region (0.3 species per 100 km^2) clearly demonstrates the biological importance of this habitat (Burgess et al. 2003). With more study, this richness continues to be uncovered: in the period 1993-2003, 4 new mammal species, two new reptile species, and a new species of amphibian were described from coastal forests (Burgess et al. 2003).

The coastal forests make a geochemical as well as biological contribution to the biosphere. Godoy et al. (2011) estimate a total of 17 Megatonnes (Mt, million tonnes) of carbon (C) are stored in the remaining coastal forest of Tanzania as of 2007, with 0.63 Mt carbon dioxide (CO_2) emitted from forest clear-cutting per year in the period 2000-2007.

The diverse systems which together make up the coastal forests are fragile, being small and often isolated, and their persistent degradation and fragmentation by the human populations that surround them imperils their continued existence (Burgess & Clarke 2000). While the general framework for drivers of anthropogenic deforestation discussed in the introduction holds true for coastal Tanzania, more precise coastal forest threats were identified in an Eastern Africa Coastal Forest workshop (Younge et al. 2002). These major threats were, in descending order of priority: inappropriate agricultural practices, charcoal burning and fuel wood extraction, wildfires, unsustainable logging, and unplanned settlements. Secondary threats were also identified, namely: poorly placed roads and infrastructure, unsustainable extraction of non-timber forest products, invasive alien species, pollution, destructive mineral extraction, and poaching (Younge et al. 2002). The root causes of these threats were felt to be poverty, lack of alternative livelihoods, and corruption in Kenya and Tanzania. Between 1990 and 2000, the coastal region of Tanzania was deforested of closed-canopy tree cover from patches >0.02 km^2 by 37 km^2 per year, equating to a rate of ~1% per year (Godoy et al. 2011). A substantial change in the accessibility of the southern half of coastal Tanzania occurred in August 2003, with the completion and opening of the Mkapa Bridge spanning the Rufiji River. At the time of its opening, this bridge was the longest of its kind in east and southern Africa. The bridge greatly increased access to some of the last remaining viable stands of woodland and coastal forest which were previously inaccessible from the north during seasonal flooding (Milledge & Kaale 2005).

During the 30 year German occupation, 26 forest reserves containing coastal forest were gazetted, which were re-declared during British rule, along with 19 new coastal forest reserves. Since independence in 1963, at least 13 new coastal forest reserves were gazetted (Burgess & Clarke 2000). The World Database on Protected Areas (WDPA) lists close to 100 protected areas in the coastal region (precise numbers depend on definition of coastal region) (IUCN & UNEP 2012).

2.1.c Study area

Figure 2.1 Map of coastal Tanzania showing the study area. The study area is shown in blue; major urban centres shown as red dots; commercial capital city, Dar es Salaam, shown as large red dot. The yellow letter 'A' (close to the top of the study area) indicates the portion covered in a previous analysis (Larrosa 2011).

2.2 REMOTE SENSING

The development of remote sensing marked a great leap forward in our understanding of global processes. Its uptake by ecologists and conservationists lagged behind those in physical geosciences, but it is now considered an important tool for mapping, observation, analysis and measurement, monitoring over space and time, and decision support (Horning et al. 2010). Remote sensing and subsequent image analysis is possibly the most frequently used technique for mapping changes in land cover (Nagendra 2008), along with field surveys which are still typically required to identify habitat types, to locate representative areas of that habitat to generate a spectral signature, and to assess the accuracy of the remote sensed data (Green et al. 2005).

As of 2012, there are 999 operational satellites in orbit (UCS 2012). Of these, around 130 are used for earth observation, earth science, remote sensing, or meteorology. Availability of data varies and Horning et al. (2010) provide a summary of satellites, sensors, and databases relevant to conservation and ecology. These range from very high resolution imagery like that provided by the IKONOS satellite, which has been used to directly observe large marine mammal populations (Abileah 2001), to the use of coarse resolution imagery to measure global tropical forest area (Mayaux et al. 1998).

Two common sources of data are medium resolution Landsat imagery, and freely available MODIS data at a lower resolution.

2.2.a **Landsat**

The Landsat programme, jointly managed by NASA and the U.S. Geological Survey, is the longest running mission of its kind, and has provided a continuous record of changes in the Earth's surface for 40 years. The Earth Resources Technology Satellite, the programme's first satellite (later renamed Landsat 1), was launched in 1972. The most recent addition to programme, Landsat 7, was successfully launched in 1999. Its predecessor, Landsat 6, failed at launch due to a ruptured fuel manifold. A Landsat Data Continuity Mission satellite is due to be launched in January of next year (Rocchio 2012).

The Landsat 7 satellite orbits at an altitude of 705 km, and has a repeat coverage period of 16 days. It carries the Enhanced Thematic Mapper Plus (ETM+), an opto-mechanical sensor with 8 bands covering the spectral range 0.45 - 12.5 μm at a resolution of 30 m (15 m for 0.52 - 0.9 μm). Scenes produced cover an area of 183 km by 170 km (Rocchio 2012). The utility of Landsat for monitoring land-cover, and specifically the detection of tropical forest clearing, has been shown in various studies (e.g. Harper et al. (2007); Christie et al. (2007); Oliveira et al. (2007)). While changes in tropical dry forest are more difficult to detect (due to varying deciduousness and understory (Tabor et al. 2010)), the technique has been demonstrated for the Cerrado in Brazil (Brannstrom et al. 2008), the Gran Chaco in Bolivia (Killeen et al. 2007), and the Atlantic forest in Paraguay (Huang et al. 2009).

2.2.b **MODIS**

The Moderate-Resolution Imaging Spectroradiometer (MODIS) is an instrument aboard both the Terra and Aqua satellites, which form part of NASA's Earth Observing System (EOS). Together, these two satellites view the entire Earth's surface every 1 to 2 days, collecting data in 36 spectral bands along a swath 2330 km by 10 km, while orbiting at the same altitude as Landsat 7 (705 km). The spatial resolution of this sensor varies by spectral band, with bands 1 (0.620 - 0.670 μm) and 2 (0.841 - 0.876 μm) at 250 m, 3 to 7 (0.459 - 2.155 μm) at 500 m, and bands 8 to 36 (0.405 μm - 14.385 μm) at 1000 m (Maccherone 2012).

The high temporal resolution of this sensor and the variety of spectral bands it senses provide insight into global dynamics and processes occurring on land, in oceans, and in the troposphere (Maccherone 2012). Many data products are derived from MODIS observations, of which two are used in this study. The MOD12Q1 Land Cover product gives global 500 m land cover based on International Geosphere-Biosphere Programme (IGBP) classifications, with a temporal granularity of 1 year and temporal coverage starting in 2001 (LPDAA 2012a). The IGBP classification system recognises 17 classes, covering natural vegetation (11 classes), developed and mosaic lands (3 classes) and non-vegetated lands (3 classes). A complete list of classes and their definitions is given by Friedl et al. (2002).

The MOD44B Vegetation Continuous Fields (VCF) product gives global 250 m sub-pixel percentage estimates of tree, herbaceous, and bare ground cover, as well as leaf type and longevity, derived from 16-day composites of reflective and emissive bands. Temporal coverage begins in 2000 (LPDAA 2012b).

Achieving accurate land cover classification for the coastal region is complex as the habitats are fragmented and heavily mosaicked, and transitions between cover types in this landscape are smooth and graduated, with no distinct boundaries (Tabor et al. 2010). While contiguous intact forest and clear-cut forest are comparatively easy to identify, the signature left by degradation and selective logging is much more difficult to detect (Tabor et al. 2010). Cloud cover, though less of a problem than in other eastern Africa coastal areas (Godoy et al. 2011), presents difficulties, obscuring large areas and casting shadow.

The supplemented Landsat data are the result of extensive validation, which incurs costs and uses resources including time. Data of similar quality are not available for all regions and times. The failure of Landsat 7's Scan Line Corrector in the ETM+ instrument on 31st May 2003 means some areas are imaged twice, and others not imaged at all (Markham et al. 2004). In contrast, MODIS data are freely available and cover the globe. A literature review found no comprehensive comparisons of supplemented Landsat ETM+ data with MODQ1 and MOD44B, though Harris et al. (2005) found MOD12Q1 imagery predicted a 20% smaller forested area than Landsat ETM+ imagery for forest patches larger than 10 km^2, and had lower agreement for smaller patches. Liu et al. (2006) showed low agreement between visually interpreted Landsat imagery and MOD44B data for estimating continuous tree distribution in China, with estimates of forest pixels from Landsat being four times higher for densely forested areas and four times lower for areas of sparse forest. The Landsat data used in these comparisons was supplemented and validated in a different manner to that used in this study, so results for this study may differ.

2.3 MODELLING DEFORESTATION

Motivations driving the modelling of deforestation are various, and the approach taken depends on these motivations, data availability, resources and scale (Lambin 1994). Confounding factors are the complexity of the system and the intricate interactions between environmental, socio-economic and cultural aspects of deforestation (Geist & Lambin 2002).

The ultimate and proximate drivers of anthropogenic deforestation vary spatially and temporally, and deforestation is typically driven by regional patterns of synergies in causal factors (Geist & Lambin 2002). These include economic factors, governmental and social institutions, and national policies at the distal level, which drive the proximate causes - namely agricultural expansion, wood extraction, and infrastructure extension (Geist & Lambin 2002). Population growth and shifting cultivation are often cited as the most prominent drivers of forest loss (Lambin et al. 2001), although Geist & Lambin (2002) suggest too much emphasis is placed on these factors as primary causes of deforestation.

Statistical modelling allows the contributions and interactions of such factors to be assessed. The most basic approach would be a linear model, with deforestation as a function of a single term: say, distance to nearest road, or slope. With an increase in this distance or slope, you might expect a proportional decrease in deforestation, as accessibility decreases. However, things are rarely this simple, as multiple factors contribute to deforestation, and these factors can interact in complex ways (Geist & Lambin 2002). A study by Sader et al. (2001) demonstrates this interaction: they found that forest on steep ground remained intact in areas far from roads, but once a road was built nearby the steepness required to confer protection on a forest patch increased.

Methods have been developed to take into account these interactions, as well as the non-linear relationship found between variables. Generalized Linear Models (GLMs) allow the response variable to have a non-normal distribution, to be discrete (e.g. binomial), and to be given by a monotonic function of the linear predictor (Wood 2006). Generalized Additive Models (GAMs) are a semi-parametric extension to GLMs, where the linear predictor is the sum of smoothing functions applied to the explanatory variables (Wood 2006; Hastie & Tibshirani 1990). Smoothing functions typically require large datasets and are computationally intensive to produce (Hastie & Tibshirani 1990; Wood 2006).

GAMs are well suited to situations where the relationship between variables is complex and not easily fitted by standard linear or nonlinear models. However, they are more complicated to fit and require a greater degree of judgment than GLMs, and the results can be harder to interpret (Wood 2006).

Given that land cover categories are typically discrete, another common approach to statistically modelling deforestation is the use of logistic regressions (Ludeke et al. 1990), which differs from linear regressions in that the response variable is binary or discrete rather than continuous (Hosmer & Lemeshow 2000). These studies model land cover change from one category to another over time. Chomitz & Gray (1996) developed a multinomial logistic regression with 3 categories for the response variable - natural vegetation, commercial agriculture, and subsistence agriculture.

A review of the literature suggests GLMs are more commonly used than GAMs, with response variables including annual rate of deforestation (Armenteras et al. 2011), probability of deforestation (McDonald & Urban 2006), and forest pattern metrics (Pan et al. 2004). Scale ranges from global (Pahari & Murai 1999) to local (Henderson-Sellers et al. 1993). Fewer studies employ GAMs to model deforestation. Mendes & Junior (2012) used a GAM to model the relationship between deforestation, corruption, and economic growth using annual deforestation rate as a response variable. Chaves et al. (2008) model the spread of disease using disease incidence rate as the response variable and deforestation as an explanatory variable. Green et al. (unpublished) use a binomial forest/non forest response variable in a GAM for modelling deforestation in the Eastern Arc Mountains.

Other modelling approaches include artificial neural networks (Pijanowski et al. 2002; Mas et al. 2004) which use a machine learning approach that has its roots in artificial intelligence research, and spatial transition-based models based on cellular automata (Theobald & Hobbs 1998).

Previous deforestation modelling efforts for Tanzania have used a binary response variable in binomial GLMs and GAMs (Larrosa 2011; Green et al. unpublished). No models with a focus on deforestation have been produced for the coastal region, and a literature review found no models using deforestation severity (defined in section 3.3.a) as the response variable.

2.3.a **Spatial autocorrelation**

Spatial autocorrelation is the correlation of a variable with itself through space. This violates the assumption of most statistical analyses that the values of data points are independent of each other. This may bias parameter estimates and increase type I error rates, where the null hypothesis is falsely rejected (Dormann et al. 2007). At a coarse resolution, forests are typically positively autocorrelated, although the degree of spatial autocorrelation varies with scale (Gilbert & Lowell 1997). At very fine scales, negative spatial autocorrelation can be seen in the non-random distribution of trees, as those large individuals that come to dominate the canopy crowd out others (Gilbert & Lowell 1997).

To measure the extent of spatial autocorrelation, and to test the assumption of independence, numerous indices have been developed, including Moran's *I* (Moran 1950), Geary's *C* (Geary 1954), Ripley's *K* (Ripley 1977), and join count analysis (Fortin & Dale 2005). While Moran's *I* is one of the oldest indicators of spatial autocorrelation, it is effective and the mostly commonly used index (Li et al. 2007). By comparing a value at a given location to the value at all other locations, it returns an index between -1 and 1, where values approaching -1 indicate strong negative correlation, 0 indicates perfect randomness of distribution, and values approaching 1 indicate strong positive correlation (i.e. a non-random clustering).

GAMs can be more robust to spatial autocorrelation than GLMs, through the inclusion of the coordinates of the grid cells as a smoothed term which, while not addressing the problem of spatial autocorrelation, can be used to account for it across large distances (Cressie 1993).

3 METHODS

3.1 STUDY AREA CHARACTERIZATION

The study area (shown in Figure 2.1) was chosen to cover the Tanzanian portion of the coastal forests of eastern Africa, following the definition given by Burgess & Clarke (2000). The precise boundaries were determined by availability of supplemented Landsat scenes (see below); modifications to these boundaries were to remove a portion covered by a previous analysis (Larrosa 2011), and to clip the images to the extent of terrestrial Tanzania. The total area enclosed by the study boundary was 81,495.54 km^2.

3.2 DATA PREPARATION

For handling and preparing spatial data I used ArcMap 9.3.1 (ESRI 2009) and Quantum GIS (Quantum GIS Development Team 2012). For statistical analysis and modelling I used R (R Core Team 2012). For reading spatial data in R, I used the 'raster' package (Hijmans & van Etten 2012).

3.2.a Land-cover change datasets

Three remotely sensed land-cover change datasets were used – supplemented Landsat data, MOD12Q1 data, and MOD44B data. The Landsat land-cover change dataset, produced by Conservation International (CI) in partnership with Sokoine University of Agriculture (SUA) and the World Wildlife Fund For Nature (WWF), comprised six Landsat ETM+ scenes (Table 3.1) supplemented with data from aerial surveys, field surveys and local knowledge. These six scenes cover 82,660 km^2 of terrestrial coastal Tanzania, and the majority of the Tanzanian Northern and Southern Zanzibar-Inhambane coastal forest mosaic, at a resolution of 28.5 m.

Table 3.1 Landsat ETM+ scenes and acquisition dates. ~2000 dates are from (Tabor et al. 2010); ~2007 dates are from (Godoy et al. 2011).

Scene		Acquisition date	
Path	*Row*	*~2000*	*~2007*
166	63	30/01/2003	09/01/2007
166	64	30/01/2003	25/11/2007
166	65	30/06/2000	29/02/2006
166	66	30/06/2000	19/05/2008
165	66	22/05/2000	10/05/2007
165	67	22/05/2000	25/03/2008

For the original dataset compiled by Tabor et al. (2010), two temporally separate images for each scene were stacked and 7 cover classes (forest, woodland, mangrove, non-forest/non-woodland, water, cloud, and cloud shadow) and 36 change classes (transitions between cover classes, where not all possible transitions occurred) were mapped with a Maximum Likelihood Classification (MLC).

The forest class was composed of mature, primary, closed canopy forest, including an area of humid tropical montane forest in the East Usambaras (Tabor et al. 2010). This area was excluded as it is found at a higher elevation than the rest of the study area, and is covered by a previous analysis (Larrosa 2011). The woodland class was composed of deciduous *Brachystegia* woodlands where crowns were adjacent, but non-overlapping. Grassland, shrubs, sparse woodland, plantations, and forest regeneration were classed as non-forest/non-woodland (Tabor et al. 2010).

The original classification scheme disaggregates woodland and forest; for this analysis, I grouped these two classes into a single class (hereafter referred to as 'forest'). This was because the processes behind their respective deforestation are likely to be similar (Neil Burgess, pers. comm.). This contrasts with the Eastern Arc Mountains where drivers have been shown to differ across these two types (Larrosa 2011; Green et al. unpublished), with differences including distance to main roads being a significant predictor of forest removal, but not of woodland – instead, distance to nearest secondary road was, suggesting woodland timber extraction is at a local scale (Larrosa 2011). Those pixels that at no point during the study period contained forest or woodland were excluded from modelling analysis (though were retained for the comparison of land-cover change datasets), as were any pixels containing cloud or cloud shadow in either time period.

With this reclassified data, I calculated the mean of an 8 by 8 pixel window (corresponding to ~240 m) and resampled this to give a proportion forest cover at a resolution of 240 m. A lower resolution was needed to a) decrease the processing load during spatial analysis, and b) calculate the continuous response variable. This exact resolution was chosen because it is approximately the native resolution of MOD44B (250 m), it includes the approximate resolution of Landsat as a factor (30 m) and it is an approximate factor of the native resolution of MOD12Q1 (500 m).

The following procedure was used to prepare both the MOD12Q1 and the MOD44B datasets. Tiles covering the study area were downloaded from NASA's Reverb ECHO site. Using MODIS Reprojection Tool, I mosaicked these to form one image, converted them from Hierarchical Data Format files to GeoTIFF, reprojected them from an integerised sinusoidal projection to UTM 37S, reclassified them to a binary raster of forest and non-forest (see appendix 7.1 for reclassification schemes), and resampled the resulting rasters to 240 m. I then clipped these to the study area, described above.

Table 3.2 MODIS tiles used in the MOD12Q1 and MOD44B datasets.

Tile		Acquisition date	
h	*v*	*~2000*	*~2007*
21	09	05/03/2000 - 06/03/2001	06/03/2007 - 05/03/ 2008
21	10	05/03/2000 - 06/03/2001	06/03/2007 - 05/03/ 2008
22	09	05/03/2000 - 06/03/2001	06/03/2007 - 05/03/ 2008
22	10	05/03/2000 - 06/03/2001	06/03/2007 - 05/03/ 2008

3.2.b Landscape characteristics dataset

Explanatory variables were selected based on an *a priori* understanding of deforestation in the region and availability of datasets, and are given in Table 3.3. These include socioeconomic factors categorised by Mitsuda & Ito (2010): accessibility (distance to roads, distance to settlements), development of local community (population density), spatial configuration (distance to forest edge, previous forest conditions) and political restriction (protection).

My protected area dataset was based on spatial data covering all protected areas in Tanzania, downloaded from the WDPA using the Protected Planet tool (IUCN & UNEP 2012). I synthesized this with data provided by WWF and SUA regarding newly gazetted protected areas, and protected areas not currently included in the WDPA dataset. While the official date of establishment of some of these protected areas is more recent than the 2000-2007 period of this study, these sites are typically proposed years before being gazetted, and their protection recognized by the surrounding communities (Neil Burgess, pers. comm.) so their effect should be detectable within the study period.

Those protected areas falling within the study area were selected for analysis. Where a protected area intersected the boundary of the study area, I included it if >50% of its area lay inside the study area. Marine parks with no terrestrial portion were excluded, as were mangrove protected areas as they contain no coastal dry forest (legal protection follows the mangrove tree line) and their boundaries are poorly described in the WDPA database (Neil Burgess, pers. comm.).

Other explanatory variables considered or initially included were distance to nearest agriculture, distance to nearest river, elevation, and slope. Due to the diffuse nature of agriculture in Tanzania and the limited resolution of the data available (300 m, GlobCover v2.3 (ESA GlobCover Project & MEDIAS-France 2009)), the distance to nearest agriculture data had an inappropriate structure for use in modelling - essentially, all pixels (save for those within some reserves and the Kilwa region) were close to agriculture, leaving the slope of

Table 3.3 Geographic, biogeographic, and socioeconomic data used as explanatory variables in deforestation severity models.

Variable name	Description	Source	Source data type	Source resolution
Continuous variables				
Distance to nearest road	Euclidean distance to nearest road in metres.	Derived from GEF project data (UNDP et al. 2010)	Polyline	-
Distance to Dar es Salaam	Distance to Tanzania's commercial capital city, Dar es Salaam.	Derived from GEF project data (UNDP et al. 2010)	Point	-
Distance to nearest urban centre	Euclidean distance to nearest urban centre (excluding Dar es Salaam), in metres.	Derived from GEF project data (UNDP et al. 2010)	Point	-
Population density	Population density - number of people per square kilometre. Settlement maps derived from satellite imagery combined with land cover maps to reallocate spatial population count data.	Tanzania AfriPop Data 2010 (Linard et al. 2012)	Raster	1 km
Latitude, longitude	The latitude and longitude of each pixel in the study. This is used to account for spatial autocorrelation across large distances.	Derived from the geocoded Landsat raster (Tabor et al. 2010)	Raster	240 m
Distance to edge of forest	Distance to nearest non-forest pixel in metres.	Derived from the Landsat data (Tabor et al. 2010)	Raster	30 m/240 m
Proportion forest cover in 2000	Continuous variable between 0 and 1, where 0 is no forest and 1 is totally forested.	Derived from the Landsat data (Tabor et al. 2010)		240 m
Categorical variables				
North or south of Rufiji River	A binary classification specifying whether a pixel is north or south of Rufiji River.	Derived from GEF project data (UNDP et al. 2010)	Polyline	-
Ethnic group	Categorical variable of main ethnic group in that location.	GREG 2010 (Weidmann et al. 2010)	Polygon	-
Region	Regions of Tanzania, cropped to the study area. Percentage of each region included given in Table 4.1.	Tanzanian National Bureau of Statistics (NBS) (2002)	Polygon	-
Protected?	A binary classification of whether a pixel falls within a protected area.	WDPA 2012 (IUCN & UNEP 2012) data updated with data derived from GEF project data (UNDP et al. 2010)	Polygon	-

the regression open to disproportionate influence from outliers. Distance to nearest river had been considered, as rivers are used to transport timber in some locations (Sokolov & Cherntsov 1995). However, this method of timber extraction is not used in Tanzania (Neil Burgess, pers. comm.). Slope and elevation have been used to predict deforestation in the Eastern Arc Mountains of Tanzania, with forest 'islands' remaining at the tops of mountains (Larrosa 2011). In the coastal region, these variables have little effect as most land is relatively flat and of a similar elevation (Neil Burgess, pers. comm.).

3.3 DATA EXPLORATION & INTERPRETATION

3.3.a Response variable

In contrast to the typical binomial models of deforestation, where a binary classification of forest/non-forest is used, here I made use of Landsat's high resolution to determine proportion forest cover within a 240 m pixel, giving an approximately continuous[1] variable, bounded at 0 (no forest) and 1 (completely forested) for 2000 and for 2007. These two layers were differenced to give a continuous index of proportion forest cover change between 2000 and 2007. Of the ~810,000 data points in this cover change variable, ~740,000 points equalled exactly 0, i.e. no change in forest cover has occurred. While this is reassuring for Tanzania's coastal forest, summarizing, exploring, and modelling the data becomes difficult, as signals and trends are swamped, and models may have difficulty producing valid results with large numbers of zeros (McCullagh & Nelder 1989). To address this problem, in this study I used the non-zero portion of the data to explore the severity of deforestation (Figure 3.1) – that is, when deforestation does occur, what proportion of the study cell is deforested? The resulting response variable varies from 0 (no deforestation of the pixel area) to -1 (complete deforestation of the pixel area) and is illustrated in Figure 3.2.

[1] Proportion forest cover for a 240 m pixel calculated from an 8x8 grid of binary 30 m pixels gives 8*8=64 possible values.

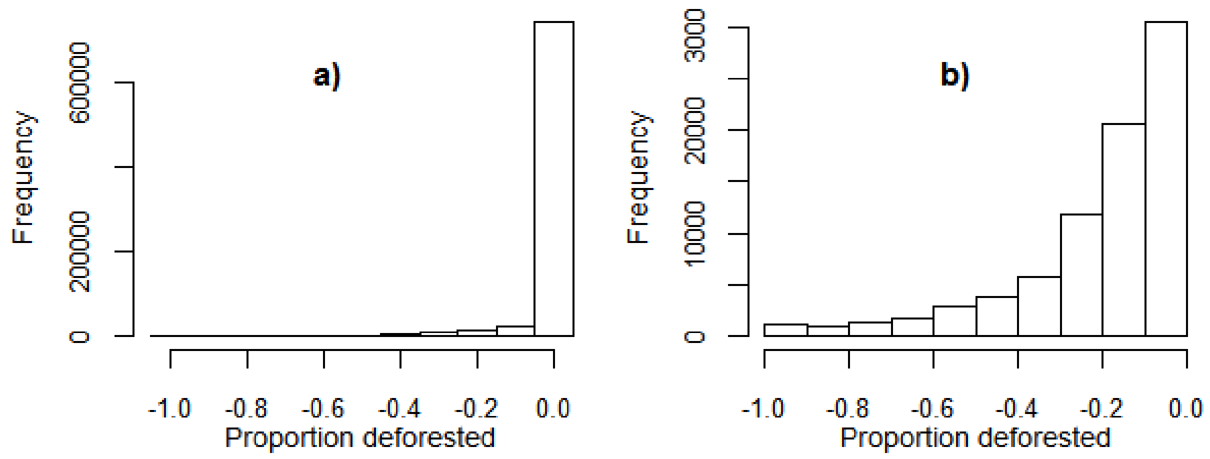

Figure 3.1 Proportion cover change including and excluding zero values. The proportion cover change shows an overwhelming signal given by the 0 values, to the extent where the distribution of the values becomes difficult to visually interpret. a) All data (-1 to 0); b) Deforestation (-1 to <0). The second plot shows the distribution of the data in the deforestation severity response variable.

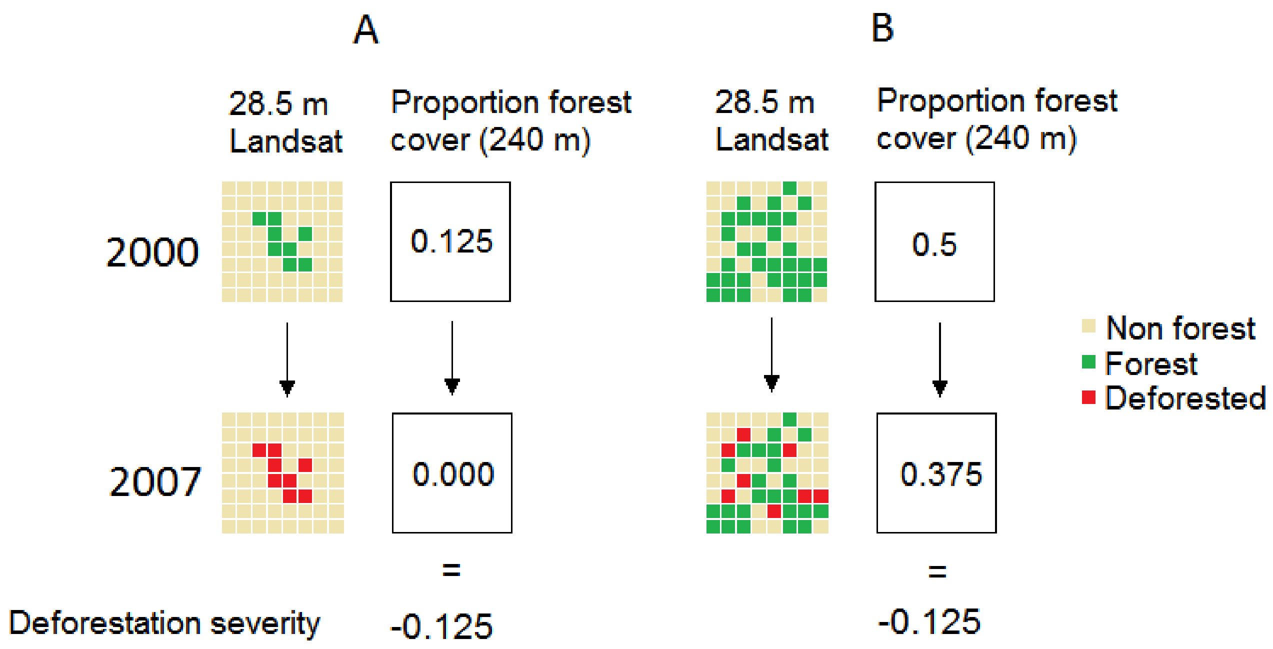

Figure 3.2 Explanation of the response variable, deforestation severity. Proportion forest cover is the number of forested pixels divided by the total number of pixels (64) in the 240 m window. Deforestation severity is the proportion cover in 2007 minus proportion forest cover in 2000. As can be seen, A and B both have the same deforestation severity. The possibility of different drivers accounting for complete clearance of small patches (seen in A) and patchy, partial clearance of large forest areas (seen in B) is discussed in section 3.6 in relation to the quartile GAMs.

For the linear model, the response variable was transformed using an arc sine square root transformation to meet assumptions of normality (McDonald 2009).

3.3.b Explanatory variables

To improve interpretability of plots, to linearize the relationship between the explanatory variable and the response for linear modelling, and to equalize the variance of the variable, I transformed the continuous explanatory variables (Box & Cox 1964). Univariate relationships were examined to determine an appropriate modelling approach (Figure 3.3 and Figure 3.4), and bivariate relationships were examined to assess independence and suitability for model inclusion, following protocol described by Zuur et al. (2010) (Figure 3.5). All continuous variables correlated pair-wise with each other by 0.6 or less, and so were retained for the model.

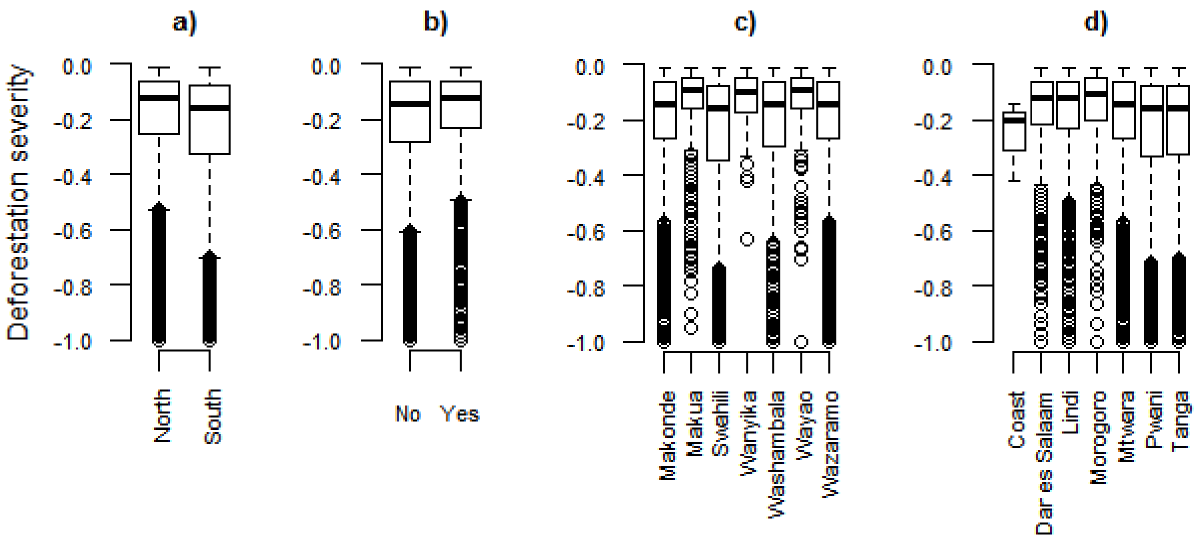

Figure 3.3 Relationship of deforestation severity with categorical explanatory variables. The deforestation severity scale runs from 0 (no deforestation of the pixel area) to -1 (complete deforestation of the pixel area). a) North or south of Rufiji River: whether a pixel lies to the north or to the south of the Rufiji river; b) Protected?: whether a pixel falls inside a protected area; c) Ethnic group: the majority ethnic group where the pixel lies; d) Region: which region of Tanzania the pixel lies in.

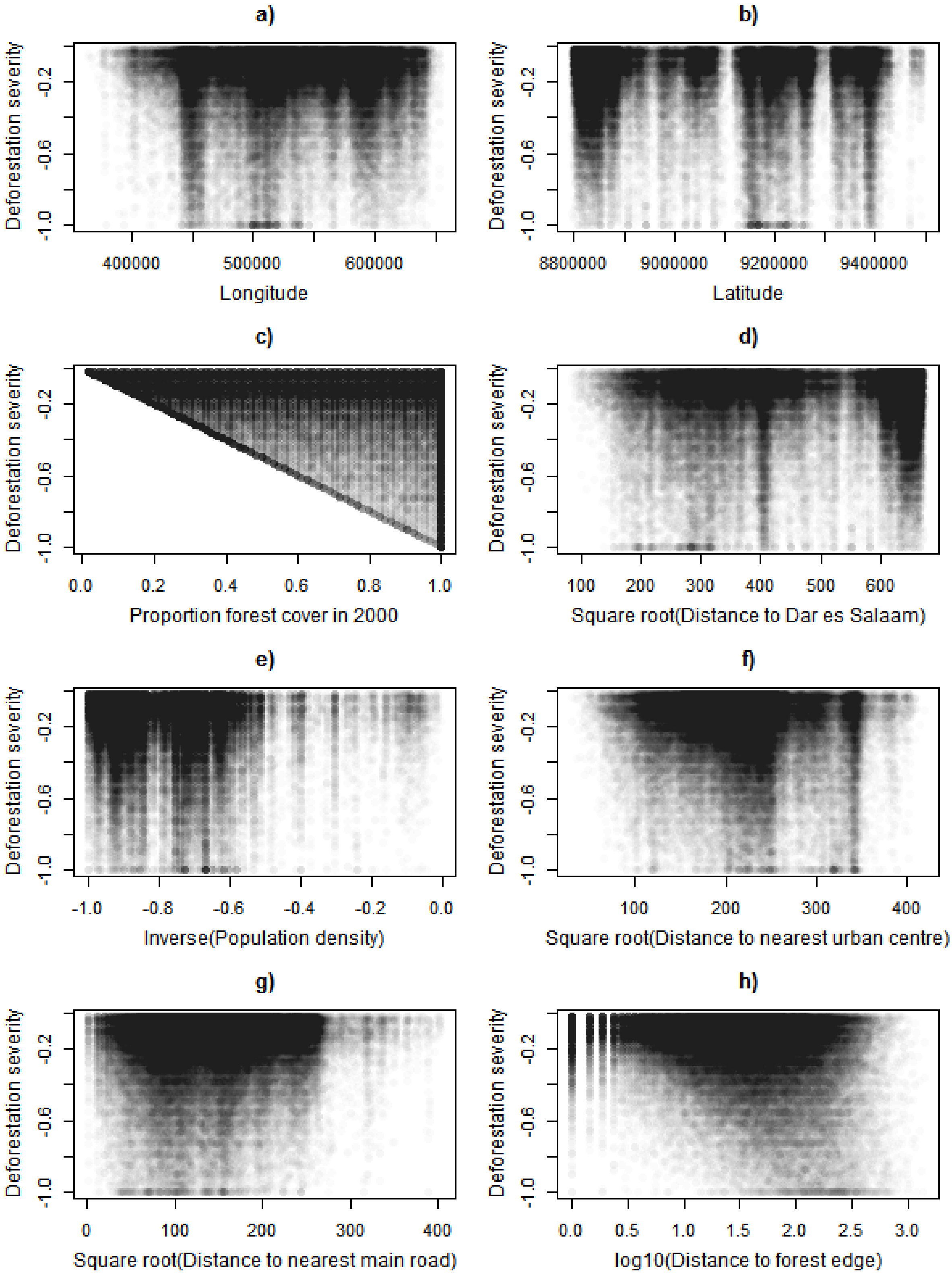

Figure 3.4 Density plots of relationship of deforestation severity with continuous explanatory variables. Darker areas correspond to more data points falling within that area, while white indicates no data points in that area. The deforestation severity scale runs from 0 (no deforestation of the pixel area) to -1 (complete deforestation of the pixel area). a) Longitude; b) Latitude; c) Proportion forest cover in 2000; d) Square root of the distance to Dar es Salaam; e) Inverse of population density; f) Square root of the distance to nearest urban centre; g) Square root of the distance to nearest main road; h) Log10 of the distance to forest edge.

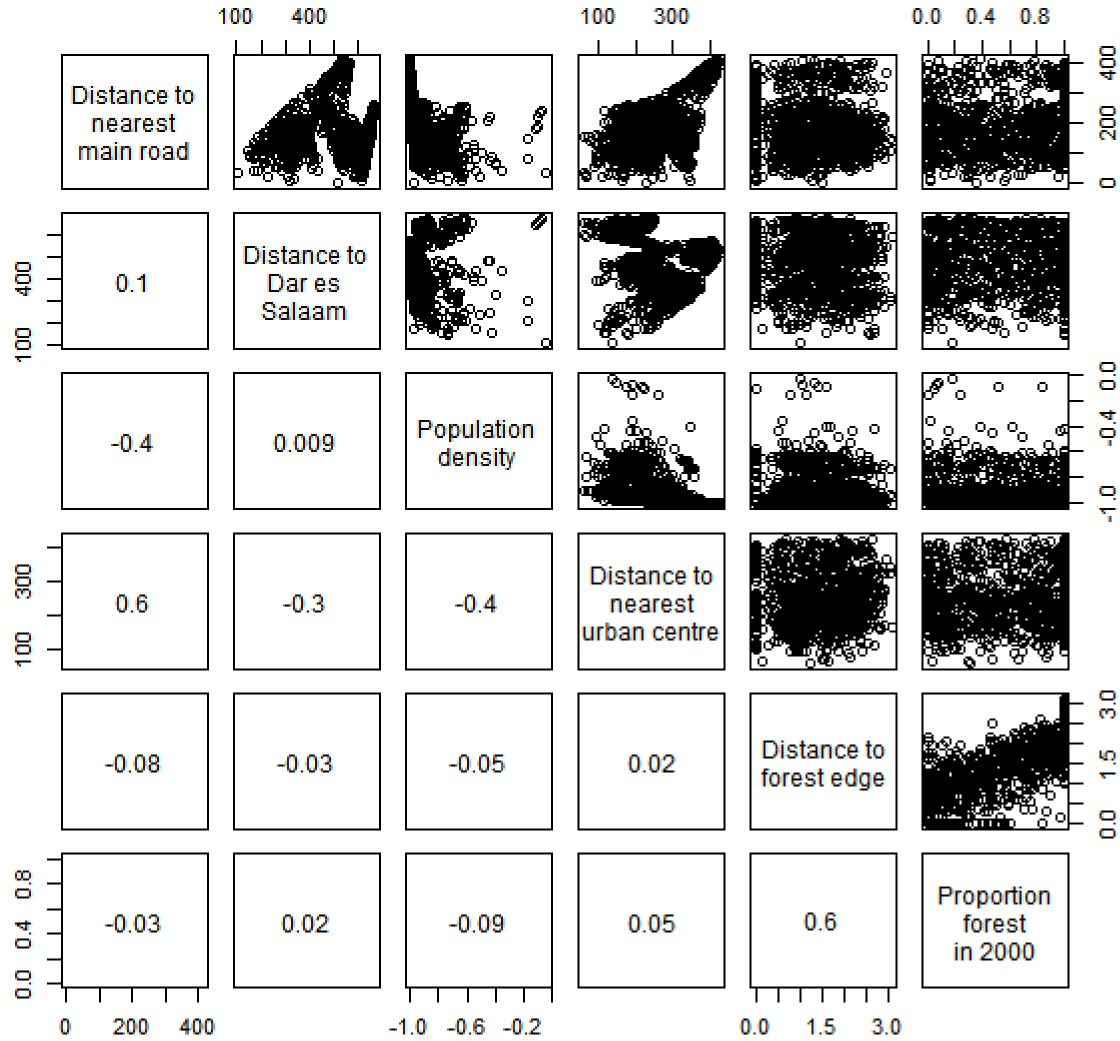

Figure 3.5 Relationships between continuous variables. The upper-right panels show pairwise scatterplots between each variable, and the lower-left panels show Pearson correlation coefficients between each variable.

3.4 QUANTIFYING DEFORESTATION

To quantify deforestation in the coastal region of Tanzania, I estimated the extent of forest in the 28.5 m supplemented Landsat data for 2000 and 2007, and calculated the difference.

3.5 COMPARISON OF LAND-COVER CHANGE DATASETS

As the MOD12Q1 data are binary, I thresholded the continuous Landsat data to a binary variable to allow comparison. A threshold of 60% forest cover was chosen to match the MOD12Q1 threshold (Strahler et al. 1999). Agreement was then observed with a contingency table, and tested using Cohen's *kappa*, a measure of inter-rater agreement which takes into account chance agreement (Cohen 1960; Carletta 1996), and an adjusted Rand index, a measure of data clustering, again taking into account chance (Rand 1971). To carry out these tests I used the 'classAgreement' function from the 'e1071' package (Dimitriadou et al. 2008) in R.

The MOD44B product gives proportion forest cover, allowing a more direct comparison with the proportion forest cover variable derived from the Landsat data. Degree of agreement was tested using linear models for the ~2000 and ~2007 scenes. Correlation was visually demonstrated with box plots for which the Landsat data were broken into 50 quantiles to allow trends to be seen in the large dataset.

3.6 DEFORESTATION MODEL

A linear model was used to explore the relationship between the response and explanatory variables. Due to the low explanatory power of this model (adjusted $R^2 = 0.185$) and the apparently nonlinear relationships seen in Figure 3.4, an approach that could capture this nonlinearity was needed.

Firstly, I produced a full GAM with the complete dataset. This used the full range of data, and included all terms given in Table 3.3. Continuous variables were modelled as smoothed terms where a curve of best fit was constructed from sections of cubic polynomial joined together at 'knots', while categorical variables were treated as factors. In an effort to increase the explanatory power of the model, I subset the data into quartiles based on the extent of forest cover at the beginning of the study period. The reasoning for this was that areas with different levels of initial forest cover would experience different forms of deforestation. This is supported by Ahrends et al. (2010), who found that high value timber is extracted first (reducing proportion cover), followed by medium value goods (again decreasing cover) and finally low value goods. The different drivers of these processes were hoped to be captured by sectioning the data into the following quantiles: first quartile model with forest cover in 2000 <25% (n =29,973); second quartile model with forest cover in 2000 ≥25% to <50% (n=23,456); third quartile model with forest cover in 2000 ≥50% to <75% (n=23,400); fourth quartile model with forest cover in 2000 ≥75% (n=44,685). For each quartile, a separate GAM was produced using the same explanatory terms.

In this analysis, I used the 'mgcv' package (Wood 2011), which includes the 'bam' function, specifically designed to fit GAMs with large datasets (n > ~20,000). To account for spatial autocorrelation, I included coordinates of the pixels as a smoothed term (Cressie 1993), reducing the spatial structuring of the model residuals.

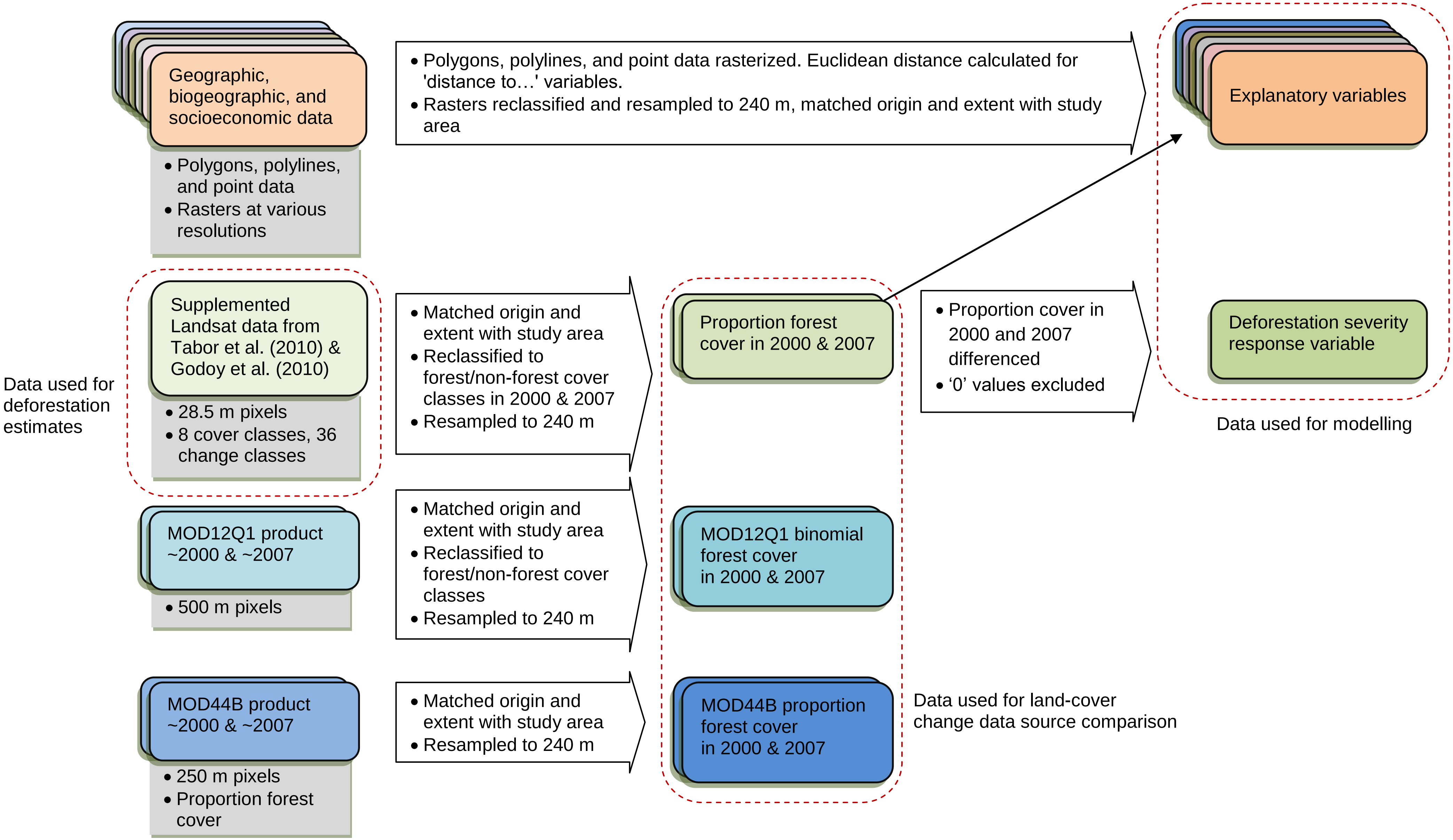

Figure 3.6 Flow diagram of data preparation methods. Coloured boxes represent data sets, white arrows indicate data processing. Black arrow indicates proportion forest cover derived from the Landsat data was included as an explanatory variable. Moving from left to right indicates increasing levels of data processing during the study. Red dotted boxes indicate sets of data used for each section of analysis. Green indicates Landsat or Landsat derived data; light blue indicates MOD12Q1 product or MOD12Q1 derived data; dark blue indicates MOD44B product or MOD44B derived data.

4 RESULTS

4.1 DEFORESTATION IN COASTAL TANZANIA 2000 - 2007

The 28.5 m supplemented Landsat data give 32,139,607 forest pixels in 2000 and 30,937,176 forest pixels in 2007. A total of 5,190,608 pixels were classified as cloud or cloud shadow across the study period. For the 81,495.54 km^2 study area, this equates to a minimum of 26,043.59 km^2 of combined forest and woodland in 2000 (32.0% of the study area), reduced to a minimum of 25,069.42 km^2 in 2007 (30.8% of the study area). Over the 7 year study period, 974.17 km^2 were deforested, giving a yearly deforestation rate of 139.17 km^2 per year. Cloud and cloud shadow covered 4,213.11 km^2 of the study area, or 5.17%. Table 4.1 shows this deforestation disaggregated by region. The regions with the highest area deforested were Pwani and Mtwara, though only 46% of Mtwara was included in the study and so likely has the highest absolute area deforested when this is accounted for. The Dar es Salaam region lost the highest percentage of its forest cover, at almost 20% across the whole study period, or an average yearly rate of 2.84%. Morogoro had the lowest total area deforested and the lowest percentage of forest cover in 2000, but only 3% of the region was included in the study area so these values may not be representative of the whole region.

Table 4.1 Deforestation in coastal Tanzania between 2000 and 2007 disaggregated by district. Due to cloud cover in the original data, these figures represent a minimum.

	% of region	**Total deforestation**		**Deforestation rate**	
District	**in study area**	*km^2*	*% of forest in 2000*	*km^2 year $^{-1}$*	*% year $^{-1}$*
Dar es Salaam	100	10.73	19.91	1.53	2.84
Lindi	52	208.68	1.56	29.81	0.22
Morogoro	3	9.21	2.42	1.32	0.35
Mtwara	46	304.75	8.68	43.54	1.24
Pwani	88	306.61	4.15	43.80	0.59
Tanga	29	134.19	10.17	19.17	1.45
Total	**-**	974.17	3.74	139.17	0.53

4.1.a Deforestation and protected areas

The process of updating the WDPA data with data from WWF and SUA resulted in a dataset containing 142 protected areas, covering a total of 11,737.8 km^2 (14% of the study area). Deforestation rates were four times higher outside protected areas than inside (Table 4.2).

Table 4.2 Deforestation in coastal Tanzania between 2000 and 2007 disaggregated by protection status. Due to cloud cover in the original data, these figures represent a minimum.

	Deforestation		
	Total (km^2)	*km^2 year $^{-1}$*	*Rate (% year $^{-1}$)*
Within protected areas	*40.75*	5.82	0.14
Outside protected areas	*933.42*	133.35	0.61

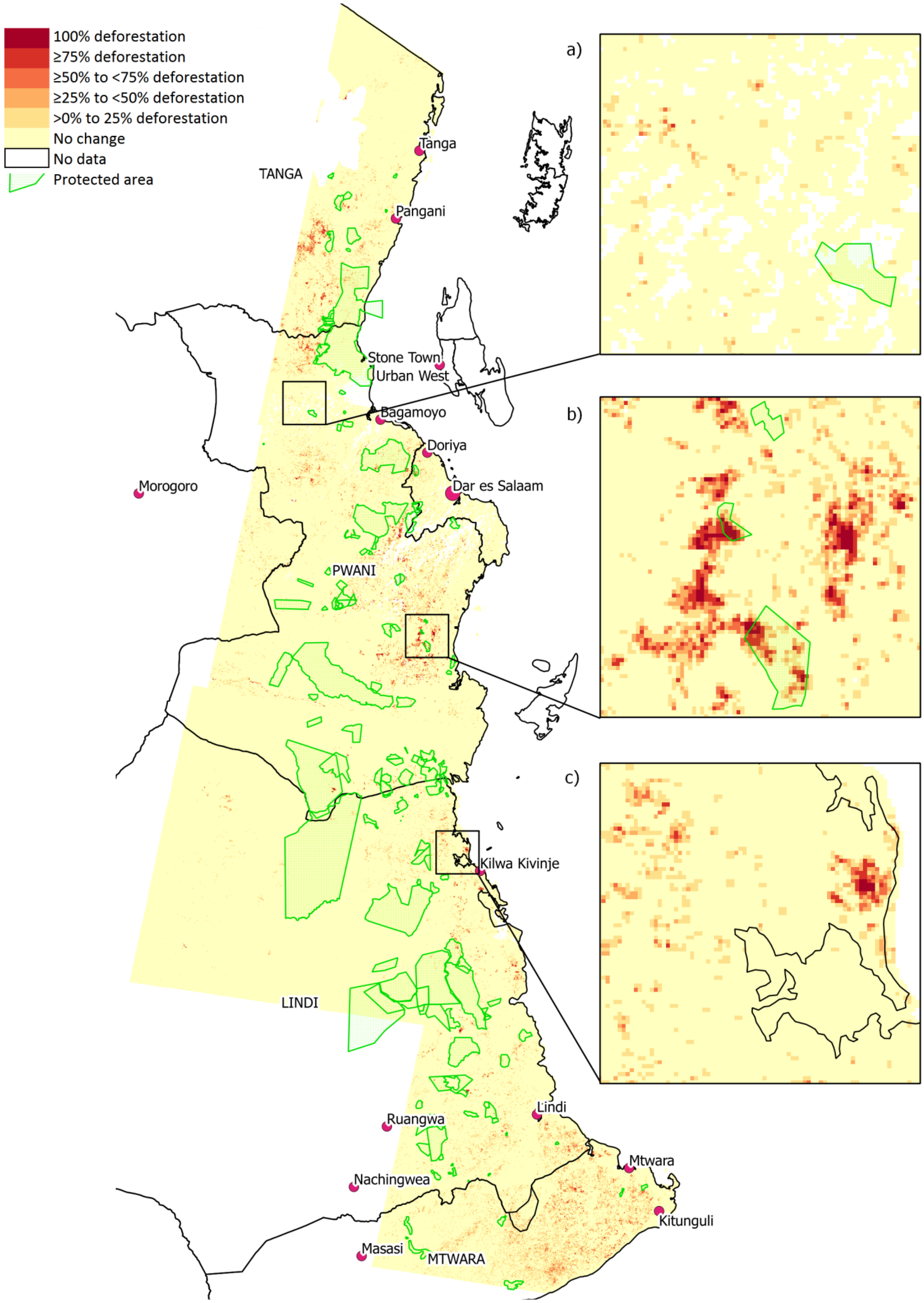

Figure 4.1 Map of deforestation in coastal Tanzania. Inset a) shows cloud cover (white pixels), insets b) and c) show severe localised inland and coastal deforestation.

4.2 COMPARISON OF LAND-COVER CHANGE DATA SOURCES

4.2.a MOD12Q1

Data for the same year from thresholded, binary Landsat and MOD12Q1 data showed reasonable percentage agreement, with more than two thirds of the data for both years falling in the diagonal of their respective contingency tables (corresponding to agreement). However, this simple assessment is biased by chance agreement. Cohen's *kappa* quantifies this chance agreement and gives a more robust index, varying between 0 and 1 where 0 indicates complete disagreement and 1 indicates complete agreement (Cohen 1960). For 2000, *kappa* for agreement between Landsat and MOD12Q1 was 0.127, and for 2007 was 0.037. These values suggest significantly lower agreement than the simple percentage scores above. This lower agreement is supported by adjusted Rand index scores of 0.06 and 0.02, which accounts for chance in the measurement data clustering, again varying between 0 and 1 for random distribution and perfect clustering respectively (Rand 1971).

Table 4.3 Contingency tables of agreement and disagreement between Landsat derived data and MOD12Q1 data. Landsat data were hardened with a threshold of 60% forest coverage to match MOD12Q1's classification scheme. Cohen's *kappa* is a measure of inter-rater agreement, and adjusted Rand index is a measure of data clustering. Both account for chance. n = 1,363,549.

		MOD12Q1 2000	
		No forest	*Forest*
Landsat 2000	*No forest*	800608	163094
	Forest	285959	113888
Percentage data agreement = 67%, Cohen's *kappa* = 0.1270, adjusted Rand index = 0.0598			

		MOD12Q1 2007	
		No forest	*Forest*
Landsat 2007	*No forest*	861066	121850
	Forest	321625	59008
Percentage data agreement = 67%, Cohen's *kappa* = 0.0370, adjusted Rand index = 0.0188			

4.2.b MOD44B

Intra-year correlations between Landsat and MOD44B are visually demonstrated in Figure 4.2. In both years, where Landsat data reports 100% forest cover, the mean MOD44B value is 30%, and there is high variability across the range of forest cover. The linear models of MOD44B as a function of Landsat data for each year give adjusted R^2 values of 0.225 and 0.239 for 2000 and 2007 respectively (P-values $<< 0.001$).

Figure 4.3 demonstrates the structural differences between Landsat and MOD44B data, with comparison between years. In Figure 4.3a, showing Landsat data, there is a slight decrease (compared to 2000 data) in values close to one in the 2007 data, and a slight increase in values close to zero. This indicates that deforestation has occurred in areas with complete forest cover in 2000, that there are a greater number of areas that have been totally deforested, and that there has been a net loss of forest during the study period. Figure 4.3b shows a very different pattern for MOD44B data, with 2007 data having a lower density of points close to zero, and a higher density of data closer to one, indicating extensive net forestation.

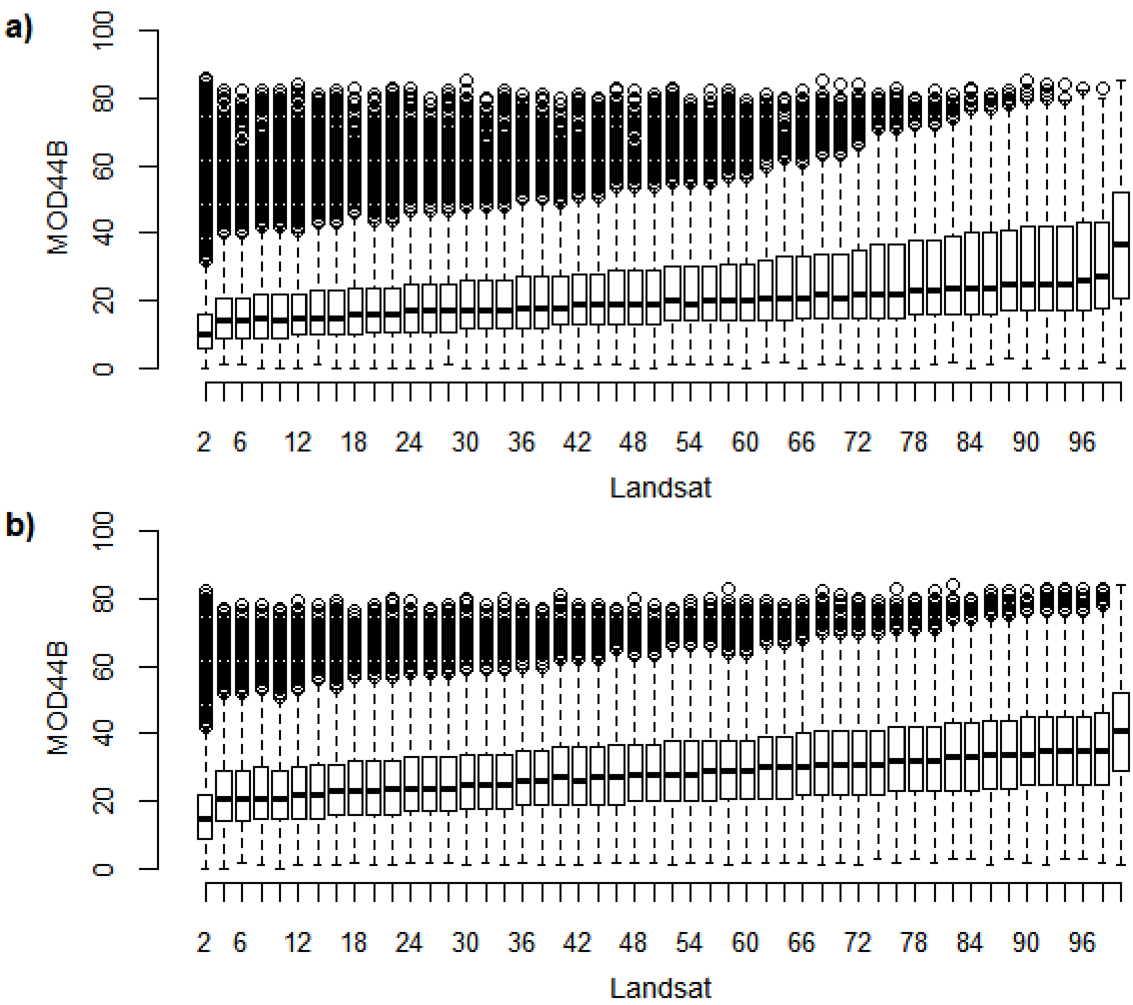

Figure 4.2 Agreement between Landsat and MOD44B data sources. a) Landsat and MOD44B data from 2000; b) Landsat and MOD44B data from 2007.

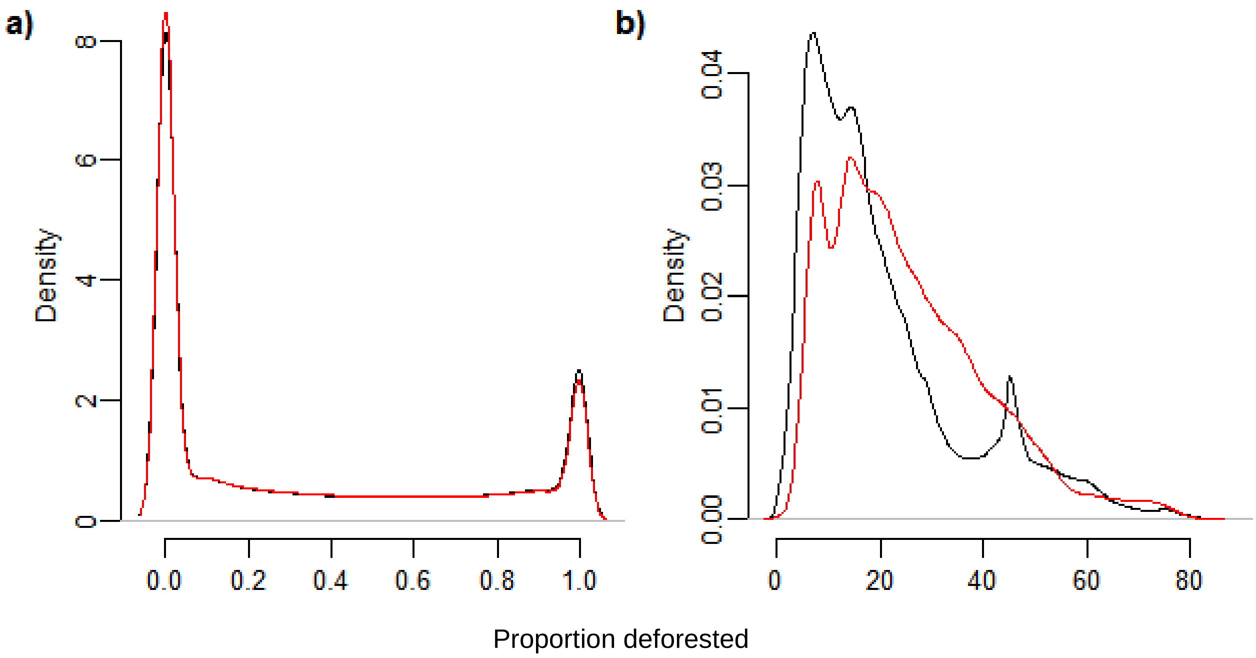

Figure 4.3 Density plots showing the structure of the Landsat and MOD44B data. The black line shows the data density for the 2000 dataset; the red shows the data density for the 2007 dataset. a) Landsat and b) MOD44B data.

4.3 PREDICTORS OF DEFORESTATION SEVERITY

4.3.a Full model

The full model used the complete data set, with all explanatory variables. All smoothed terms were highly significant. For the categorical variables, whether a pixel was protected gave a significant effect, as did the ethnic group (though only for some groups within the variable). Whether a pixel was north or south of the Rufiji River made a slight but significant difference, though the effect of this term's inclusion in the model on the Akaike Information Criterion (AIC) was positive - i.e. the model performed marginally better in its absence. The deviance explained by the full model was 20.8%. The two terms with the highest explanatory power were the proportion of forest cover in 2000, and latitude & longitude, with all other terms contributing less than 1% to the deviance explained (Table 4.4).

Table 4.4 Explanatory variable contributions to the full model. The individual contribution of terms was assessed by the iterative deletion of each variable from the model and the calculation of the percentage decrease of deviance explained and the increased AIC of the resulting model. Smoothed terms do not provide an estimate - the equivalent of an estimate would be the function describing the shape of the smoothed line. Full model: deviance explained = 20.8%, Akaike Information Criterion (AIC) =-44,766.07, n = 121,514.

Explanatory variable	Estimate	Degrees of freedom (⋄ = estimated)	Δ Deviance explained	Δ AIC	P value
Smoothed terms					
Proportion forest cover in 2000		7.7⋄	8.70%	-8391.98	<<0.001 ***
Latitude, Longitude		26.5⋄	1.10%	-1004.82	<<0.001 ***
Population density		7.7⋄	0.50%	-470.57	<<0.001 ***
Distance to forest edge		7.2⋄	0.20%	-171.91	<<0.001 ***
Distance to Dar es Salaam			0.30%	-256.54	
North of Rufiji River		6.0⋄			<<0.001 ***
South of Rufiji River		5.9⋄			0.00122 **
Distance to nearest main road		8.2⋄	0.20%	-195.33	<<0.001 ***
Distance to nearest urban centre		8.0⋄	0.10%	-83.27	<<0.001 ***
Categorical variables					
Ethnic group		6	0.10%	-61.45	
Makua	-0.012849				0.10144
Swahili	0.096979				<<0.001 ***
Wanyika	0.110251				0.00321 **
Washambala	0.083467				0.0011 **
Wayao	0.048835				0.04535 *
Wazaramo	0.132095				<<0.001 ***
Protected?	0.021150	1	0.10%	-70.73	<<0.001 ***
Region		6	0.10%	-48.74	
Dar es Salaam	0.094805				0.37876
Lindi	0.130689				0.21939
Morogoro	0.041833				0.69854
Mtwara	0.100189				0.34577
Pwani	0.093865				0.38138
Tanga	0.119153				0.26896
North or south of Rufiji River	-0.734620	1	0.00%	1.6	0.04026 *

Smooth functions differ for the distance to Dar es Salaam depending on whether a point is north or south of the Rufiji River, with the northern portion showing more variation. The relationship between these two terms and deforestation severity is shown in Figure 4.4, which also shows a decrease in deforestation severity as latitude and longitude increase (i.e. towards the north east).

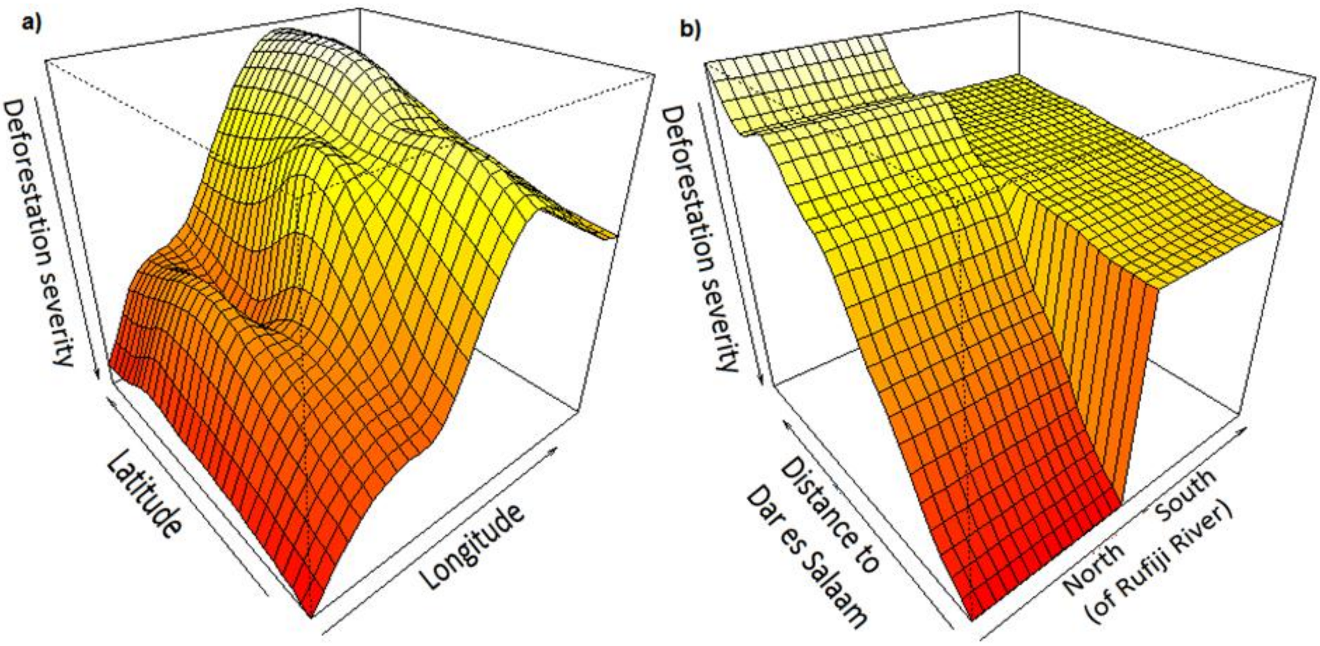

Figure 4.4 Interactions between selected variables in the full model. Red cells indicate high severity deforestation, white cells indicate low severity deforestation. a) Interaction between latitude, longitude, and deforestation severity; b) Interaction between distance to Dar es Salaam, whether a pixel is north or south of Rufiji River, and deforestation severity.

Figure 4.5 shows the component smooth functions for the continuous variables, and the two dimensional smooth for the latitude & longitude term. In Figure 4.5c, deforestation severity can be seen to increase with proportion forest cover in 2000, suggesting greater threat of deforestation for more intact forest. This trend is reflected in Figure 4.5e, where the log10 of distance to forest edge shows an initially flat relationship with deforestation severity, until it drops sharply, suggesting core forest away from the fringes of the forest patch is susceptible to severe deforestation. Figure 4.5d shows deforestation severity increasing with the inverse of population density, i.e. areas with high population density are less likely to be deforested than areas with low population density. Figure 4.5f describing the smoothed term for square root of distance to nearest urban centre shows an increase in deforestation severity as the distance increases, followed by a sharp decline with widening confidence intervals.

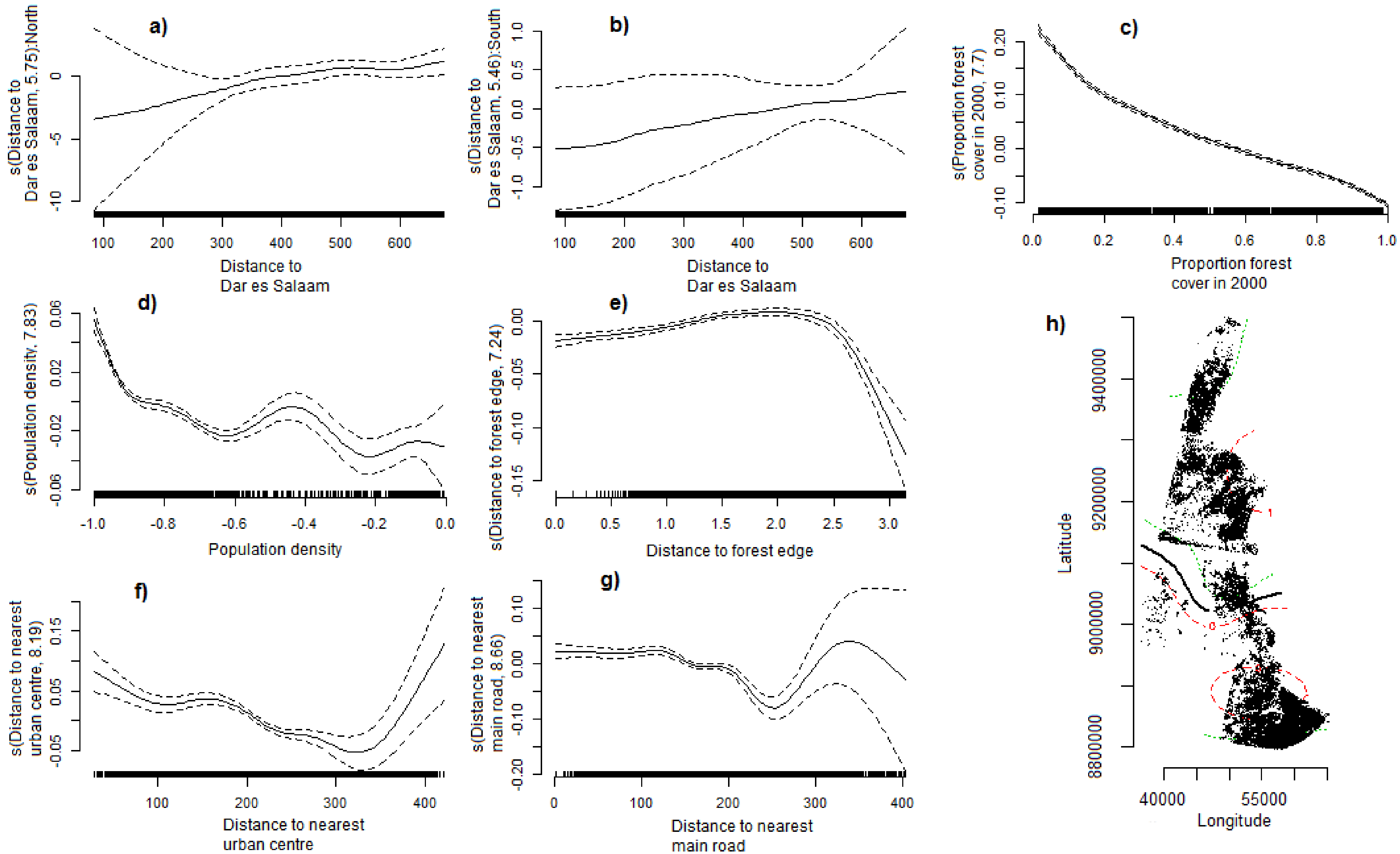

Figure 4.5 Component smooth functions of the full model. The fitted surface is shown by the black line. Dotted lines show 2 standard errors above and below the estimate. The vertical dashes along the x axis show location of data points. The number that follows the variable name on the y axis is the estimated degrees of freedom. Note the scale of the y axis differs between plots, and that lower values indicate more severe deforestation. a) Distance to Dar es Salaam, north of Rufiji River; b) Distance to Dar es Salaam, south of Rufiji River; c) Proportion forest cover in 2000; d) Population density; e) Distance to forest edge; f) Distance to nearest urban centre; g) Distance to nearest main road; h) Latitude & longitude. The red and green lines are surfaces at +1 and -1 standard errors, contoured and overlaid on the contour plot for the estimate.

4.3.b Quartile GAMs

The data used in the full model were partitioned into quartiles to capture the effect of varying proportion forest cover in 2000. This partitioning did not increase the explanatory power of the models in the manner expected. Only the first quartile model performed better than the full model (Table 4.5). Much of the explanatory power of the model came from the proportion forest cover in 2000 term, which explained 40.5% of the deviance of the complete model. The next largest contribution to the model was from the distance to forest edge term (deviance explained: 0.8%). The second, third, and fourth quartile models all performed worse than the full model, each explaining only around 11% of the deviance. For the second quartile model, the terms with the largest explanatory power were proportion forest cover in 2000, and latitude & longitude. In the third quartile model, latitude & longitude explained the most deviance, followed by proportion forest cover in 2000. For the fourth quartile model, latitude & longitude explain the most deviance, followed by the distance to Dar es Salaam and the categorical variable of whether the pixel is north or south of the Rufiji River, both explaining 0.7% of the deviance. Whether a pixel is protected or not contributes only a small amount to the model, suggesting a limited impact of protection on deforestation severity.

Table 4.5 Model summaries for quartile GAMs. The individual contribution of terms was assessed by the iterative deletion of each variable from the model and the calculation of the percentage decrease of deviance explained. Quartiles based on forest cover in 2000. First quartile <25% forest cover in 2000 (n =14,782); second quartile ≥25% and <50% (n=15,684); third quartile ≥50% and <75% (n=16,855); fourth quartile ≥75% (n=33,742). See Appendix 7.2 for full model plots and checking plots.

	First quartile model	Second quartile model	Third quartile model	Fourth quartile model
Explanatory variable	*Deviance explained*			
Smoothed terms				
Proportion forest cover in 2000	40.5%	3.0%	0.9%	0.7%
Latitude, Longitude	0.6%	1.7%	1.8%	1.6%
Population density	0.1%	0.4%	0.5%	0.6%
Distance to forest edge	0.8%	0.4%	0.2%	0.1%
Distance to Dar es Salaam, *by North or south of Rufiji River*	0.1%	0.3%	0.4%	0.7%
Distance to nearest main road	0.1%	0.4%	0.2%	0.0%
Distance to nearest urban centre	0.1%	-0.1%	0.2%	0.1%
Categorical variables				
Ethnicity	0.1%	0.1%	0.3%	0.0%
Protected?	0.0%	0.0%	0.2%	-0.1%
Region	0.0%	0.1%	0.1%	0.0%
North or south of Rufiji River	0.1%	0.3%	0.4%	0.7%
Deviance explained by full quartile model	42.50%	12.50%	10.30%	11.80%

4.3.c Spatial autocorrelation

All variables showed highly significant autocorrelation. 'Distance to…' variables showed the highest degree of spatial autocorrelation, and severity of deforestation showed the lowest (Table 4.6).

Table 4.6 Moran's *I* for each continuous variable. *Observed* is the computed Moran's *I*; *Expected* is the expected value of *I* under the null hypothesis that there is no spatial autocorrelation; Standard Deviation is the standard deviation of *I* under this hypothesis. All *P*-values <<0.001. Due to computational load, the test was run on a random subsample of the data (n=8,000).

	Moran's *I*		**Standard**
Variable name	*Expected*	*Observed*	**deviation**
Distance to Dar es Salaam	-0.0002	0.744991	0.000932
Distance to nearest urban centre	-0.0002	0.403127	0.000932
Distance to nearest main road	-0.0002	0.152674	0.000932
Population density	-0.0002	0.098217	0.000931
Proportion forest cover in 2000	-0.0002	0.078867	0.000932
Distance to forest edge	-0.0002	0.072946	0.000932
Severity of deforestation	-0.0002	0.045580	0.000932

5 DISCUSSION

5.1 DEFORESTATION IN COASTAL TANZANIA 2000 – 2007

The rate of deforestation found for the study area, 0.53% per year, is higher than that reported by Godoy et al. (2011), who found a rate of loss of 0.4% per year for the coastal region[2]. The difference is likely to be due to the grouping of woodland and forest in this study, and suggests that rate of loss is higher in woodland than forest. This is supported by the higher rates of woodland loss in the Eastern Arc Mountains reported by Larrosa (2011), who found an annual rate of deforestation for this region of 1.55% from 1970 to 2000. The FAO reports forest was lost across Tanzania at a rate of 1.14% per year between 2000 and 2007 (FRA 2010). This is the second highest percentage rate in mainland eastern and southern Africa, behind Uganda with a rate of loss of 2.39% per year. It should be noted that these data are self-reported, and that a new evaluation based on remote-sensed data suggest overestimation of both forest extent and rate of loss in the report (FAO 2011).

It appears that coastal forests experience a comparatively low rate of deforestation, at least in absolute terms. The pattern of deforestation may be more important than the magnitude, and increasing fragmentation and degradation could still imperil species and ecosystems.

5.1.a Deforestation and protected areas

It is reassuring to find the percentage deforestation rates were 4 times lower inside protected areas than outside (0.14% per year and 0.61% per year respectively). Larrosa (2011) found a difference between protected and unprotected forest in the same direction, though of a smaller magnitude, with an annual deforestation rate of 0.77% inside protected areas and 2.19% outside protected areas (2.8 times lower inside than outside). However, a simple analysis of deforestation rates inside and outside protected areas is a crude measure of their success (Andam et al. 2008). Grouping all protected areas into one category disguises trends and patterns. Protected area designation varies, with some having the explicit goal of preventing deforestation, while others are intended for use (Newmark et al. 1993). Within protected areas with this latter designation, forest cover may fluctuate, especially over the comparatively short study period of 7 years. Such an analysis also does not take into account spillover – i.e. increased rates of deforestation in the immediate vicinity of a protected area

---

[2] It should be noted that these figures both represent a minimum due to cloud cover in the original remotely sensed data.

due to the displacement of deforestation from inside its boundaries (Andam et al. 2008), which has been shown to occur in eastern Africa's evergreen forests (Pfeifer et al. 2012). This approach also fails to account for the non-random placement of protected areas, which are often designated in places unsuitable for other uses due to remoteness or high elevation (Joppa & Pfaff 2010).

The protected area data themselves, while the best available for the project and checked for obvious mistakes, may contain spatial errors like those found in the data covering the Eastern Arc Mountains (Larrosa 2011). These include incorrect boundaries, spatial shifts, and crude approximates of protected area shape. This could be corrected by redigitising the boundaries from maps in parallel with expert opinion and alignment with obvious spatial features such as rivers or distinctive forest patch shape, though this would have been beyond the scope of this project and may have taken several months to complete (Lauren Coad, pers. comm.).

5.2 COMPARISON OF LAND-COVER CHANGE DATA SOURCES

5.2.a MOD12Q1

Comparisons between Landsat and MOD12Q1 data show 67% agreement. While this seems high, it should be noted that chance agreement does not account for 50% as might be assumed (Wood 2007). For a 10 by 10 grid of pixels where only one cell is forested, random predictions of the values in cells will agree 99% of the time on the classification of the unforested areas, even if they disagree on the location of the forested cell. In the same way, predictions of forest distribution can show high percentage agreement while disagreeing on the location of much of the forest, through the agreement for the majority cover class, non-forest. Cohen's *kappa* and the adjusted Rand index give more robust accounts of agreement. To place the values for agreement between Landsat and MOD12Q1 (for 2000, *kappa* = 0.127, for 2007, *kappa* = 0.037) in context, agreement of psychiatric and medical diagnoses averages *kappa* values of between 0.60 and 0.70 (Wood 2007).

Harris et al. (2005) found MOD12Q1 imagery predicted a smaller total forested area than Landsat ETM+ imagery and had low agreement for patches smaller than 10 km^2. The ability of Landsat to detected smaller patches with more accuracy than MOD12Q1 may account for the low agreement between these two sources in this study.

5.2.b MOD44B

While a positive correlation can be seen between Landsat and MOD44B data within each year, the trend is weak with high variability across the range of forest cover. Linear models of

MOD44B data as a function of Landsat data show significant moderate correlation, with R^2 of above 0.20 for both years. This is in line with Liu et al. (2006), who showed a similarly low agreement between visually interpreted Landsat imagery and MOD44B data for estimating continuous tree distribution in China.

The correlation found here may be misleading; while intra-year comparisons show correlation in the same direction, inter-year comparisons show significant differences. The apparent forestation is refuted by the more robust Landsat data and by expert ground-truthing (Neil Burgess, pers. comm.).

Complete agreement between data sources would not be expected, as they use different cover classes defined by different criteria. For this study, the Landsat data were taken as the authority, but even in this validated dataset, errors may be present. As with all optical satellite imagery used for land cover mapping, the Landsat data were affected by cloud cover, as well as the challenges of defining vegetation classes in a heavily mosaicked, transitional, and complex landscape. Transitions between cover types in this landscape are smooth and graduated, with no distinct boundaries. While clear-cut forest is comparatively easy to identify, the signature left by selective logging is much more difficult to detect (Tabor et al. 2010).

These results show that the supplemented Landsat product is significantly different from the two MODIS products, MOD12Q1 and MOD44B, supporting the need for the labour intensive process of validation and ground-truthing. Agreement between Landsat and MOD12Q1 might be improved with the use of different classification schemes, but the MOD44B product appears unsuitable due to the very different inter-year signal. NASA notes that caution should be exercised when making inter-annual comparisons with the MOD44B product and recommends that, until a multi-year change product is derived, change can be detected in the MOD44A product, Vegetative Cover Conversions (Carroll et al. 2006). However, this product would have been less suited to comparison with the continuous deforestation severity response variable derived from the Landsat data as it gives binary (change/no change) classes.

Inappropriate use of these MODIS datasets could lead to the under-reporting of deforestation, as the loss of smaller forest patches and degradation may be missed. Inaccurate quantification of forest and forest loss may result in the misallocation of funding in programmes such as REDD+, and invalid conclusions may be drawn about whether CBD targets have been met.

5.3 MODEL PERFORMANCE

5.3.a Full model

Increasing proportion forest cover in 2000 predicted more severe deforestation by 2007. This suggests that intact forest is most at threat, and that previously degraded forest is comparatively safe, perhaps because it has been drained of valuable products. This effect was explored further in the quartile models, and is discussed further in the following section. Latitude & longitude explained the second largest amount of deviance. This term is discussed further in section 5.3.d.

Population density in this model explained only a small amount of deviance, and is a poor predictor of deforestation severity. In binomial deforestation models for the Eastern Arc Mountains (Larrosa 2011) and other areas (Carr 2004), population density is often a highly significant predictor (though the correlation can be positive or negative depending on the situation (Carr 2004)). This suggests that, while population density can be used to predict where deforestation occurs, it is less able to predict the severity of that deforestation.

With the completion of Mkapa Bridge over the Rufiji River in 2003, areas of previously inaccessible forest were opened up (Milledge & Kaale 2005). With increased access for loggers it would seem likely that deforestation patterns would homogenise across what was previously a barrier to timber extraction. However, there remains a significant difference between the area north and the area south of Rufiji River, perhaps due to a time lag.

Whether a pixel is protected or not only contributes a small amount to the model. This may seem surprising, especially in the context of the higher rate of deforestation found outside protected areas than inside. However, it should be remembered this is a model of deforestation *severity* and explains only the extent of deforestation when and where deforestation does occur. A more correct interpretation of this result is that protected areas, while having a lower deforestation rate, do not show a substantial difference in deforestation severity compared to outside protected areas. Protected areas in the coastal region are not respected in terms of charcoal production (Neil Burgess, pers. comm.). This result may suggest that, when an area is deforested for charcoal production, its status as protected does not cause less complete clearing, with only a few trees taken here and there. Rather, the pattern of use is the same inside as outside the protected area.

The explanatory power of the full GAM was not as high as found in similar GAMs developed for the Eastern Arc Mountains (Green et al. unpublished). An explanation considered for this was the relationship of the response variable with proportion forest cover in 2000. As can been in Figure 3.2, complete clearance of sparse forest, and the partial clearance of dense forest can have the same value in the deforestation severity response variable. These two processes could differ in their drivers. For example, Patch B could have the valuable trees removed from the area, while change in Patch A could be the removal of remnant trees to make way for agriculture. A wave-like action of forest degradation has been demonstrated in Tanzania, with the removal of high value products from comparatively intact forest being followed by the removal of degraded forest for lower value products, the drivers and correlates of which differ (Ahrends et al. 2010). This temporal process means that much of the forest around the commercial capital, Dar es Salaam, was already removed prior to 2000 (Godoy et al. 2011), preventing its inclusion in this model. This may have influenced the relationship of population density and deforestation, which showed a decrease in deforestation severity in areas of high population density. The year 2000 does not represent the start of the deforestation process, and areas close to Dar es Salaam (with high population density) were likely drained of valuable products many years ago, as the initial wave of extraction passed through, followed sequentially by the extraction of lower value goods. The gradual increase in deforestation severity followed by the sharp decline above a certain distance may show the front of this wave of deforestation emanating from Dar es Salaam, highlighting areas likely to experience deforestation in the near future.

The exclusion of any agricultural terms is likely to have reduced the performance of the model. Miyamoto (2010) found the main driver of tropical deforestation to be agricultural expansion. The GlobCover dataset was initially considered for inclusion in the model but, as discussed in section 3.2.b, without considerable refinement it would have been inappropriate for inclusion in the model. Other data are available (from Tanzania's National Bureau of Statistics (NBS), for example) but their preparation would have been beyond the scope of this short project. Numerous other explanatory variables may be used, incorporating economic data such as wage rates, agricultural prices, household income, and tenure security (Kaimowitz & Angelsen 1998).

The processing time needed to produce the GAMs using the 'mgcv' package (Wood 2011), and particularly the 'bam' function, was significantly shorter than expected, with the full model (n=121,514) taking less than a minute to be calculated. For a future analysis, a higher resolution dataset may be used to achieve more precise results.

5.3.b Quartile models

The effect of proportion forest cover in 2000 becomes clearer in the quartile models. While the relationship remains fairly stable for the second, third, and fourth quartile GAMs, proportion forest cover explains a large amount of deviance for the first quartile model. The smooth function of this term shows a steeper slope (appendix 7.2) than in other models, and suggests that areas with smaller forest cover remain untouched compared to areas with slightly higher forest cover. An explanation for this may be that these fragments have been drained of all valuable products, and are made up of trees with no commercial value in areas unsuited to other uses - perhaps with low soil fertility (Chomitz & Gray 1996). While better than complete deforestation, these small fragments are unlikely to provide much of a refuge for large vertebrates. They may, however, provide habitat for a number of forest species that can persist for long periods of time in small fragments (Turner & Corlett, 1996).

An alternative explanation for this could be the structure of the data rather than any causal link. Figure 3.4c shows the relationship between deforestation severity and proportion forest cover in 2000. If a pixel's proportion forest cover in 2000 is 20%, the greatest amount of forest that can be removed is 20% of the pixel's area. This gives rise to the shape seen in Figure 3.4c. If a line of best fit was plotted through this data, and the data were divided up into four quartiles along the x axis, the average of the residuals for the first quartile would be much lower than the average of the residuals of the fourth quartile. In the model this would be expressed as higher explanatory power, when the tighter correlation is actually due to the constrained structure of the data rather than a causal link.

5.3.c Response variable

While the use of the proportional response variable, essentially a measure of deforestation severity, superficially seems a more precise measure than a forest/non-forest binary response, as it captures information at a sub-pixel level, the model performance was lower than a similar (though binomial) model for the Eastern Arc Mountains (Green et al. unpublished). An interesting next step for this project would be to derive a binomial forest change response variable from the same data and produce a GAM to compare with the continuous severity response variable.

The grouping of woodland and forest may have impacted model performance. Distinguishing woodland from farmed landscapes with planted trees is difficult, and this classification has a lower accuracy than that for forest (Godoy et al. 2011).

5.3.d Spatial autocorrelation

The high level of spatial autocorrelation in the 'distance to…' variables is explained by the fact these variables are essentially gradients from one or several focal points or lines, with the number and distribution of these focal points or lines affecting the degree of spatial autocorrelation. The spatial autocorrelation in the distance to forest edge variable and the proportion forest cover in 2000 variable can be explained by the underlying spatial autocorrelation in the non-random distribution of the forest in the coastal region. It is interesting to note the comparatively low spatial autocorrelation for deforestation severity. This means that in a local area where one degree of deforestation severity has occurred, it is only marginally more probable that the surrounding area will be deforested at the same severity. While the inclusion of the latitude & longitude term accounts for spatial autocorrelation, this implementation does not allow the effects of spatial autocorrelation to be separated from environmental effects correlated geographically (e.g. climate gradients towards the equator) (Elith & Leathwick 2009).

As Legendre (1993) notes, natural systems without spatial structuring would be unlikely to function. In ecological studies, it is often not a case of whether spatial autocorrelation is present, but to what extent (Kissling & Carl 2008). Dormann et al. (2007) review other methods to account for spatial autocorrelation which include autocovariate regression, conditional and simultaneous autoregressive models, and spatial eigenvector mapping, and Carl & Kühn (2008) propose a wavelet-based method to remove spatial autocorrelation in species distribution models, an approach that may be applicable to the modelling of deforestation.

5.4 CLOUD COVER

The remotely sensed forest cover and change estimates are confounded by frequent cloud cover in this area (Godoy et al. 2011) combined with the low temporal resolution (Xiao et al. 2009). Percentage rates of deforestation should be indicative of trends for the whole area, while absolute rates are a minimum.

5.5 RECOMMENDATIONS

With further work, the use of a deforestation severity response variable may provide a different insight into deforestation processes than binomial forest/non forest response variables. The use of this approach was partly an attempt to capture sub-pixel information when resampling to reduce computational load. However, with the development of tools able to calculate models using very large datasets in a short period of time, the need to capture sub-pixel information at a low resolution may be reduced - instead, a higher resolution dataset may be used in its entirety (avoiding problems of representativeness when using random samples of a dataset).

This study has demonstrated the differences between supplemented, labour intensive datasets and freely available datasets for coastal Tanzania, and supports the continued production and use of supplemented, validated datasets.

Further work will be needed to produce models of deforestation in the coastal region with the explanatory and predictive power to accurately assess the effectiveness of protected areas. Inclusion of appropriate agricultural data may have a large impact on model performance.

5.6 CONCLUSION

This thesis describes the first model of deforestation in coastal Tanzania, and the only example of a model using deforestation severity. This response variable has potential future use in modelling, providing different insights into deforestation processes. While the model produced has limitations, it has highlighted trends and patterns of deforestation in a fragile and imperilled landscape. Further work relating deforestation to protected areas will improve the protection afforded this threatened ecosystem.

6 REFERENCES

Abileah, R., 2001. Use of high resolution space imagery to monitor the abundance, distribution, and migration patterns of marine mammal populations. In *MTS/IEEE Conference and Exhibition OCEANS, 2001*. MTS/IEEE Conference and Exhibition OCEANS, 2001. pp. 1381 –1387 vol.3.

Achard, F., 2006. Remote sensing and data availability. Available at: http://unfccc.int/files/methods_and_science/lulucf/application/pdf/060830_achard.pdf [Accessed February 9, 2012].

Acosta, I., 2011. "Green desert" monoculture forests spreading in Africa and South America. *The Guardian*. Available at: http://www.guardian.co.uk/environment/2011/sep/26/monoculture-forests-africa-south-america [Accessed July 26, 2012].

Adams, W.M., 2004. *Against Extinction: The Story of Conservation*, Earthscan.

Ahrends, A. et al., 2010. Predictable waves of sequential forest degradation and biodiversity loss spreading from an African city. *Proceedings of the National Academy of Sciences*. Available at: http://www.pnas.org/content/early/2010/07/28/0914471107 [Accessed August 13, 2012].

Andam, K.S. et al., 2008. Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the National Academy of Sciences*, 105(42), p.16089.

Armenteras, D. et al., 2011. Understanding deforestation in montane and lowland forests of the Colombian Andes. *Regional Environmental Change*, 11(3), pp.693–705.

Barnosky, A.D. et al., 2011. Has the Earth's sixth mass extinction already arrived? *Nature*, 471(7336), pp.51–57.

Boden, T.A., Marland, G. & Andres, R.J., 2010. Global, Regional, and National Fossil-Fuel CO2 Emissions. *Carbon Dioxide Information Analysis Center*. Available at: http://cdiac.ornl.gov/trends/emis/tre_glob.html [Accessed June 7, 2012].

Box, G.E.P. & Cox, D.R., 1964. An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), pp.211–252.

Bradshaw, C.J.A. et al., 2007. Global evidence that deforestation amplifies flood risk and severity in the developing world. *Global Change Biology*, 13, pp.2379–2395.

Brannstrom, C. et al., 2008. Land change in the Brazilian Savanna (Cerrado), 1986–2002: Comparative analysis and implications for land-use policy. *Land Use Policy*, 25(4), pp.579–595.

Brooks, T.M. et al., 2002. Habitat loss and extinction in the hotspots of biodiversity. *Conservation biology*, 16(4), pp.909–923.

Burgess, N.D. et al., 2006. Factoring species, non-species values and threats into biodiversity prioritisation across the ecoregions of Africa and its islands. *Biological Conservation*, 127(4), pp.383–401.

Burgess, N.D. et al., 2010. Getting ready for REDD+ in Tanzania: a case study of progress and challenges. *Oryx*, 44(03), pp.339–351.

Burgess, N.D. et al., 2003. New information on the lowland coastal forests of eastern Africa. *Oryx*, 37, pp.280–281.

Burgess, N.D. & Clarke, G.P., 2000. *Coastal Forests of Eastern Africa*, IUCN.

Carl, G. & Kühn, I., 2008. Analyzing spatial ecological data using linear regression and wavelet analysis. *Stochastic Environmental Research and Risk Assessment*, 22(3), pp.315–324.

Carletta, J., 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, 22(2), pp.249–254.

Carr, D.L., 2004. Proximate Population Factors and Deforestation in Tropical Agricultural Frontiers. *Population and environment*, 25(6), pp.585–612.

Carroll, M.L. et al., 2006. Vegetative Cover Conversion MOD44A. *Burned Vegetation, Collection*, 4.

Castree, N., 2010. Neoliberalism and the Biophysical Environment 1: What "Neoliberalism" is, and What Difference Nature Makes to it. *Geography Compass*, 4(12), pp.1725–1733.

CBD, 2010. Strategic Plan for Biodiversity 2011–2020 and the Aichi Targets. Available at: http://www.cbd.int/doc/strategic-plan/2011-2020/Aichi-Targets-EN.pdf [Accessed August 23, 2012].

Chape, S. et al., 2005. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1454), pp.443–455.

Chaves, L.F. et al., 2008. Social Exclusion Modifies Climate and Deforestation Impacts on a Vector-Borne Disease. *PLoS Negl Trop Dis*, 2(2), p.e176.

Chomitz, K.M. & Gray, D.A., 1996. Roads, Land Use, and Deforestation: A Spatial Model Applied to Belize. *The World Bank Economic Review*, 10(3), pp.487–512.

Christie, T. et al., 2007. Fragmentation and Clearance of Liberia's Forests During 1986-2000. *Oryx*, 41(04), pp.539–543.

CIA, 2012. CIA - The World Factbook - Tanzania. Available at: https://www.cia.gov/library/publications/the-world-factbook/geos/tz.html#.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), pp.37–46.

Cressie, N., 1993. *Statistics for Spatial Data* Revised ed., Wiley-Interscience.

Dávalos, L.M. et al., 2011. Forests and Drugs: Coca-Driven Deforestation in Tropical Biodiversity Hotspots. *Environmental Science & Technology*, 45(4), pp.1219–1227.

Dimitriadou, E. et al., 2008. *Misc functions of the Department of Statistics (e1071), TU Wien*, Available at: cran. r-project. org/web/packages/e1071/index. html [Accessed August 27, 2012].

Dirzo, R. & Raven, P., 2003. Global state of biodiversity and loss. *Annu. Rev. Environ. Resour.*, 28, pp.137–167.

Dormann, C. et al., 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, 30(5), pp.609–628.

Elith, J. & Leathwick, J.R., 2009. Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), pp.677–697.

Eltahir, E.A.B. & Bras, R.L., 1994. Sensitivity of regional climate to deforestation in the Amazon basin. *Advances in Water Resources*, 17(1–2), pp.101–115.

ESA GlobCover Project & MEDIAS-France, 2009. Global Land Cover v2.3. Available at: http://ionia1.esrin.esa.int.

ESRI, 2009. *ArcMap*, Redlands, California: Environmental Systems Resource Institute.

FAO, 2011. Satellite technology yields new forest loss estimates. *FAO Media Centre:* Available at: http://www.fao.org/news/story/en/item/95180/icode/ [Accessed August 29, 2012].

Fortin, M.-J. & Dale, M.R.T., 2005. *Spatial Analysis: A Guide For Ecologists*, Cambridge University Press.

FRA, 2010. *Global Forest Resources Assessment 2010: Main Report*, Food and Agriculture Organization of the United Nations.

Friedl, M.A. et al., 2002. Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, 83(1-2), pp.287–302.

Geary, R.C., 1954. The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, 5(3), pp.115–146.

Geist, H.J. & Lambin, E.F., 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation. *BioScience*, 52(2), pp.143–150.

Gilbert, B. & Lowell, K., 1997. Forest attributes and spatial autocorrelation and interpolation: effects of alternative sampling schemata in the boreal forest. *Landscape and Urban Planning*, 37(3–4), pp.235–244.

Godoy, F.L. et al., 2011. Deforestation and CO2 emissions in coastal Tanzania from 1990 to 2007. *Environmental Conservation*, 39(01), pp.62–71.

Goldstein, G., 2004. Legal System and Wildlife Conservation: History and the Law's Effect on Indigenous People and Community Conservation in Tanzania, The. *Georgetown International Environmental Law Review*, 17, p.481.

Green, E.P. et al., 2005. Remote sensing handbook for tropical coastal management. Available at: http://www.citeulike.org/group/342/article/431327 [Accessed August 29, 2012].

Green, J. et al., unpublished. An East-African perspective: predictors of deforestation in a biodiversity hotspot.

Gullison, R.E. et al., 2007. Tropical Forests and Climate Policy. *Science*, 316(5827), pp.985–986.

Harper, G.J. et al., 2007. Fifty Years of Deforestation and Forest Fragmentation in Madagascar. *Environmental Conservation*, 34(04), pp.325–333.

Harris, G.M., Jenkins, C.N. & Pimm, S.L., 2005. Refining Biodiversity Conservation Priorities. *Conservation Biology*, 19(6), pp.1957–1968.

Hastie, T.J. & Tibshirani, R.J., 1990. *Generalized Additive Models*, Taylor & Francis.

Henderson-Sellers, A. et al., 1993. Tropical Deforestation: Modeling Local- to Regional-Scale Climate Change. *Journal of Geophysical Research*, 98(D4), pp.7289–7315.

Hijmans, R.J. & van Etten, J., 2012. *raster: Geographic analysis and modeling with raster data*, Available at: http://CRAN.R-project.org/package=raster.

Horning, N. et al., 2010. *Remote Sensing for Ecology and Conservation: A Handbook of Techniques*, Oxford University Press, USA.

Hosmer, D.W. & Lemeshow, S., 2000. *Applied Logistic Regression*, John Wiley & Sons.

Huang, C. et al., 2009. Assessment of Paraguay's forest cover change using Landsat observations. *Global and Planetary Change*, 67(1–2), pp.1–12.

IPCC, 2007. *Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press.

IUCN, 2011. The IUCN Red List of Threatened Species. Version 2011.2. Available at: http://www.iucnredlist.org [Accessed June 6, 2012].

IUCN & UNEP, 2012. The World Database on Protected Areas (WDPA). Available at: www.protectedplanet.net.

Jenkins, M., 2007. *USDA Proposes Market-Based Conservation*, Washington, D.C. Available at: http://www.fs.fed.us/news/2007/releases/04/market-based-conservation.shtml [Accessed June 7, 2012].

Joppa, L. & Pfaff, A., 2010. Reassessing the forest impacts of protection. *Annals of the New York Academy of Sciences*, 1185(1), pp.135–149.

Kaimowitz, D. & Angelsen, A., 1998. *Economic models of tropical deforestation: a review*, Cifor. Available at: http://www.cifor.org/publications/pdf_files/Books/Model.pdf [Accessed August 15, 2012].

Killeen, T.J. et al., 2007. Thirty years of land-cover change in Bolivia. *Ambio*, 36(7), pp.600–606.

Kissling, W.D. & Carl, G., 2008. Spatial autocorrelation and the selection of simultaneous autoregressive models. *Global Ecology and Biogeography*, 17(1), pp.59–71.

Lambin, E.F., 1994. *Modelling deforestation processes: a review*, European Space Agency ESRIN Earthnet Programme: Office for Official Publications of the European Community.

Lambin, E.F. et al., 2001. The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change*, 11(4), pp.261–269.

Larrosa, C., 2011. *The effectiveness of a range of protected area governance and management regimes in reducing deforestation: the Eastern Arc Mountains of Tanzania*.

Legendre, P., 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology*, 74(6), pp.1659–1673.

Li, H., Calder, C.A. & Cressie, N., 2007. Beyond Moran's I: Testing for Spatial Dependence Based on the Spatial Autoregressive Model. *Geographical Analysis*, 39(4), pp.357–375.

Linard, C. et al., 2012. Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. *PloS one*, 7(2), p.e31743.

Liu, R. et al., 2006. Continuous tree distribution in China: A comparison of two estimates from Moderate-Resolution Imaging Spectroradiometer and Landsat data. *Journal of Geophysical Research*, 111(D8), p.D08101.

LPDAA, 2012a. MCD12Q1. *Land Cover Type Yearly L3 Global 500 m SIN Grid*. Available at: https://lpdaac.usgs.gov/products/modis_products_table/mcd12q1 [Accessed August 20, 2012].

LPDAA, 2012b. MOD44B. *Vegetation Continuous Fields Yearly L3 Global 250m*. Available at: https://lpdaac.usgs.gov/products/modis_products_table/mod44b [Accessed August 20, 2012].

Ludeke, A.K., Maggio, R.C. & Reid, L.M., 1990. An analysis of anthropogenic deforestation using logistic regression and GIS. *Journal of Environmental Management*, 31(3), pp.247–259.

Maccherone, B., 2012. MODIS Website. Available at: http://modis.gsfc.nasa.gov/ [Accessed July 15, 2012].

Markham, B.L. et al., 2004. Landsat sensor performance: history and current status. *IEEE Transactions on Geoscience and Remote Sensing*, 42(12), pp.2691 – 2694.

Mas, J.F. et al., 2004. Modelling deforestation using GIS and artificial neural networks. *Environmental Modelling & Software*, 19(5), pp.461–471.

Mayaux, P. et al., 1998. Global tropical forest area measurements derived from coarse resolution satellite imagery: a comparison with other approaches. *Environmental Conservation*, 25(1), pp.37–52.

McCullagh, P. & Nelder, J.A., 1989. *Generalized linear models*, Chapman & Hall/CRC.

McDonald, J.H., 2009. *Handbook of biological statistics*, University of Delaware: Sparky House Publishing Baltimore, MD. Available at: ftp://84.237.21.152/pub_archive/lin/yu/dev_books/HandbookBioStat.pdf [Accessed August 24, 2012].

McDonald, R.I. & Urban, D.L., 2006. Spatially varying rules of landscape change: lessons from a case study. *Landscape and Urban Planning*, 74(1), pp.7–20.

Mendes, C.M. & Junior, S.P., 2012. Deforestation, economic growth and corruption: a nonparametric analysis on the case of Amazon forest. *Applied Economics Letters*, 19(13), pp.1285–1291.

Meyfroidt, P., Rudel, T.K. & Lambin, E.F., 2010. Forest transitions, trade, and the global displacement of land use. *Proceedings of the National Academy of Sciences*. Available at: http://www.pnas.org/content/early/2010/11/05/1014773107 [Accessed July 26, 2012].

Milledge, S.A.H. & Kaale, B., 2005. Bridging the Gap: Linking Timber Trade with Infrastructural Development in Southern Tanzania-Baseline data before completion of the Mkapa Bridge. *TRAFFIC East/Southern Africa, Dar es Salaam, Tanzania. ISBN*, 958402590.

Mitsuda, Y. & Ito, S., 2010. A review of spatial-explicit factors determining spatial distribution of land use/land-use change. *Landscape and Ecological Engineering*, 7(1), pp.117–125.

Mittermeier, R.A. et al., 2005. *Hotspots Revisited: Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions*, Conservation International.

Miyamoto, M., 2010. The Causes of Tropical Deforestation: *Journal of the Japanese Forest Society*, 92(4), pp.226–234.

Moran, P.A.P., 1950. Notes on Continuous Stochastic Phenomena. *Biometrika*, 37(1/2), pp.17–23.

Nagendra, H., 2008. Do Parks Work? Impact of Protected Areas on Land Cover Clearing. *AMBIO: A Journal of the Human Environment*, 37(5), pp.330–337.

Newmark, W.D. et al., 1993. Conservation attitudes of local people living adjacent to five protected areas in Tanzania. *Biological Conservation*, 63(2), pp.177–183.

Nichols, J.D. & Williams, B.K., 2006. Monitoring for conservation. *Trends in Ecology & Evolution*, 21(12), pp.668–673.

Nobre, C.A., Sellers, P.J. & Shukla, J., 1991. Amazonian Deforestation and Regional Climate Change. *Journal of Climate*, 4(10), pp.957–988.

Oliveira, P.J.C. et al., 2007. Land-Use Allocation Protects the Peruvian Amazon. *Science*, 317(5842), pp.1233–1236.

Olson, D.M. et al., 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, 51(11), pp.933–938.

Pahari, K. & Murai, S., 1999. Modelling for prediction of global deforestation based on the growth of human population. *ISPRS journal of photogrammetry and remote sensing*, 54(5), pp.317–324.

Pallangyo, D.M., 2007. Environmental Law In Tanzania: How Far Have We Gone? *Law, Environment and Development Journal*.

Pan, W.K.Y. et al., 2004. Farm-level models of spatial patterns of land use and land cover dynamics in the Ecuadorian Amazon. *Agriculture Ecosystems & Environment*, 101(2-3), pp.117–134.

Pfeifer, M. et al., 2012. Protected Areas: Mixed Success in Conserving East Africa's Evergreen Forests. *PLoS ONE*, 7(6), p.e39337.

Pijanowski, B.C. et al., 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. *Computers, Environment and Urban Systems*, 26(6), pp.553–575.

Pimm, S.L. & Raven, P., 2000. Biodiversity: Extinction by numbers. *Nature*, 403(6772), pp.843–845.

Quantum GIS Development Team, 2012. *Quantum GIS Geographic Information System*, Open Source Geospatial Foundation Project. Available at: http://qgis.osgeo.org.

R Core Team, 2012. *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R-project.org.

Rademaekers, K. et al., 2010. Study on the Evolution of Some Deforestation Drivers and Their Potential Impacts on the Costs of an Avoiding Deforestation Scheme. *European Commission: Final Report*.

Rand, W.M., 1971. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336), pp.846–850.

Ripley, B.D., 1977. Modelling Spatial Patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(2), pp.172–212.

Rocchio, L., 2012. NASA - Landsat Science. Available at: http://landsat.gsfc.nasa.gov/ [Accessed July 15, 2012].

Sader, S.A. et al., 2001. Forest change monitoring of a remote biosphere reserve. *International Journal of Remote Sensing*, 22(10), pp.1937–1950.

Santilli, M. et al., 2005. Tropical deforestation and the Kyoto Protocol. *Climatic Change*, 71(3), pp.267–276.

Scharlemann, J.P.W. et al., 2010. Securing tropical forest carbon: the contribution of protected areas to REDD. *Oryx*, 44(03), pp.352–357.

Schneider, L.C. & Pontius Jr., R.G., 2001. Modeling land-use change in the Ipswich watershed, Massachusetts, USA. *Agriculture, Ecosystems & Environment*, 85(1–3), pp.83–94.

Soares-Filho, B.S. et al., 2006. Modelling conservation in the Amazon basin. *Nature*, 440(7083), pp.520–523.

Sokolov, K.B. & Cherntsov, V.A., 1995. Water transport of wood: possibilities and reality. *Lesnaya Promyshlennost'*, (3).

Strahler, A. et al., 1999. MODIS Land Cover Product Algorithm Theoretical Basis Document (ATBD) MODIS Land Cover and Land-Cover Change. *Change*, (May).

Tabor, K. et al., 2010. Forest and Woodland Cover and Change in Coastal Tanzania and Kenya, 1990 to 2000. *Journal of East African Natural History*, 99(1), pp.19–45.

Theobald, D.M. & Hobbs, N.T., 1998. Forecasting rural land-use change: a comparison of regression-and spatial transition-based models. *Geographical and Environmental Modelling*, 2, pp.65–82.

Tucker, C.J. & Townshend, J.R.G., 2000. Strategies for monitoring tropical deforestation using satellite data. *International Journal of Remote Sensing*, 21(6-7), pp.1461–1471.

Turner, I.M. & Corlett, R.T., 1996. The conservation value of small, isolated fragments of lowland tropical rain forest. *Trends in Ecology & Evolution*, 11(8), pp.330–333.

UCS, 2012. UCS Satellite Database. *Union of Concerned Scientists*. Available at: http://www.ucsusa.org/nuclear_weapons_and_global_security/space_weapons/technical_issues/ucs-satellite-database.html [Accessed July 26, 2012].

UNDP, Tanzania National Parks & Ministry of Natural Resources and Tourism, 2010. Strengthening the Protected Area Network in Southern Tanzania: Improving the Effectiveness of National Parks in Addressing Threats to Biodiversity. Available at: http://www.thegef.org/gef/node/4008.

Veit, P.G. & Benson, C., 2004. When Parks and People Collide. *Human Rights Dialogue: "Environmental Rights."* Available at: http://www.carnegiecouncil.org/viewMedia.php?prmTemplateID/8/prmID/4449.

Weidmann, N.B., Rød, J.K. & Cederman, L.-E., 2010. Representing ethnic groups in space: A new dataset. *Journal of Peace Research*, 47(4), pp.491–499.

Van der Werf, G.R. et al., 2009. CO2 emissions from forest loss. *Nature Geoscience*, 2(11), pp.737–738.

Wertz-Kanounnikoff, S. & Angelsen, A., 2009. Global and national REDD+ architecture. *Realising REDD*, p.13.

White, F., 1983. The vegetation of Africa: A descriptive memoir to accompany the UNESCO/AETFAT/UNSO vegetation map of Africa (Natural Resources Research 20). *UN Educational, Scientific and Cultural Organization, Paris*.

Wood, J.M., 2007. Understanding and Computing Cohen's Kappa: A Tutorial. *WebPsychEmpiricist. Web Journal at http://wpe. info/*. Available at: http://works.bepress.com/james_wood/22/ [Accessed August 25, 2012].

Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1), pp.3–36.

Wood, S.N., 2006. *Generalized Additive Models: An Introduction with* R, Taylor & Francis.

Xiao, X. et al., 2009. A Simple Algorithm for Large-Scale Mapping of Evergreen Forests in Tropical America, Africa and Asia. *Remote Sensing*, 1(3), pp.355–374.

Younge, A., Negussie, G. & Burgess, N.D., 2002. Eastern Africa Coastal Forest Programme: Regional Workshop Report. Available at: http://oceandocs.org/handle/1834/307 [Accessed August 2, 2012].

Zuur, A.F., Ieno, E.N. & Elphick, C.S., 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1(1), pp.3–14.

7 APPENDICES

7.1 RASTER RECLASSIFICATION SCHEMES

Table 7.1 Reclassification scheme for Landsat data.

Original Landsat land cover product	Reclassed Landsat 2000	Reclassed Landsat 2007
0	0	0
11	1	1
12	1	0
15	NoData	NoData
16	NoData	NoData
22	0	0
24	0	0
25	NoData	NoData
26	NoData	NoData
42	0	0
44	0	0
45	NoData	NoData
46	NoData	NoData
47	0	0
51	NoData	NoData
52	NoData	NoData
54	NoData	NoData
55	NoData	NoData
56	NoData	NoData
57	NoData	NoData
58	NoData	NoData
61	NoData	NoData
62	NoData	NoData
64	NoData	NoData
65	NoData	NoData
66	NoData	NoData
67	NoData	NoData
68	NoData	NoData
72	0	0
74	0	0
75	NoData	NoData
76	NoData	NoData
77	0	0
82	1	0
85	NoData	NoData
86	NoData	NoData
88	1	1
255	NoData	NoData
NoData	NoData	NoData

Table 7.2 Reclassification scheme for MOD12Q1 data.

MOD12Q1 classification	Description	Land cover in 2000/2007
1	Evergreen needleleaf forest	1
2	Evergreen broadleaf forest	1
3	Deciduous needleleaf forest	1
4	Deciduous broadleaf forest	1
5	Mixed forest	1
6	Closed shrublands	1
7	Open shrublands	1
8	Woody savannas	1
9	Savannas	0
10	Grasslands	0
11	Permanent wetlands	0
12	Croplands	0
13	Urban and built-up	0
14	Cropland mosaics	0
15	Snow/Ice	0
16	Barren or sparsely vegetated	0
17	Water bodies	0

7.2 MODEL SUMMARIES AND PLOTS

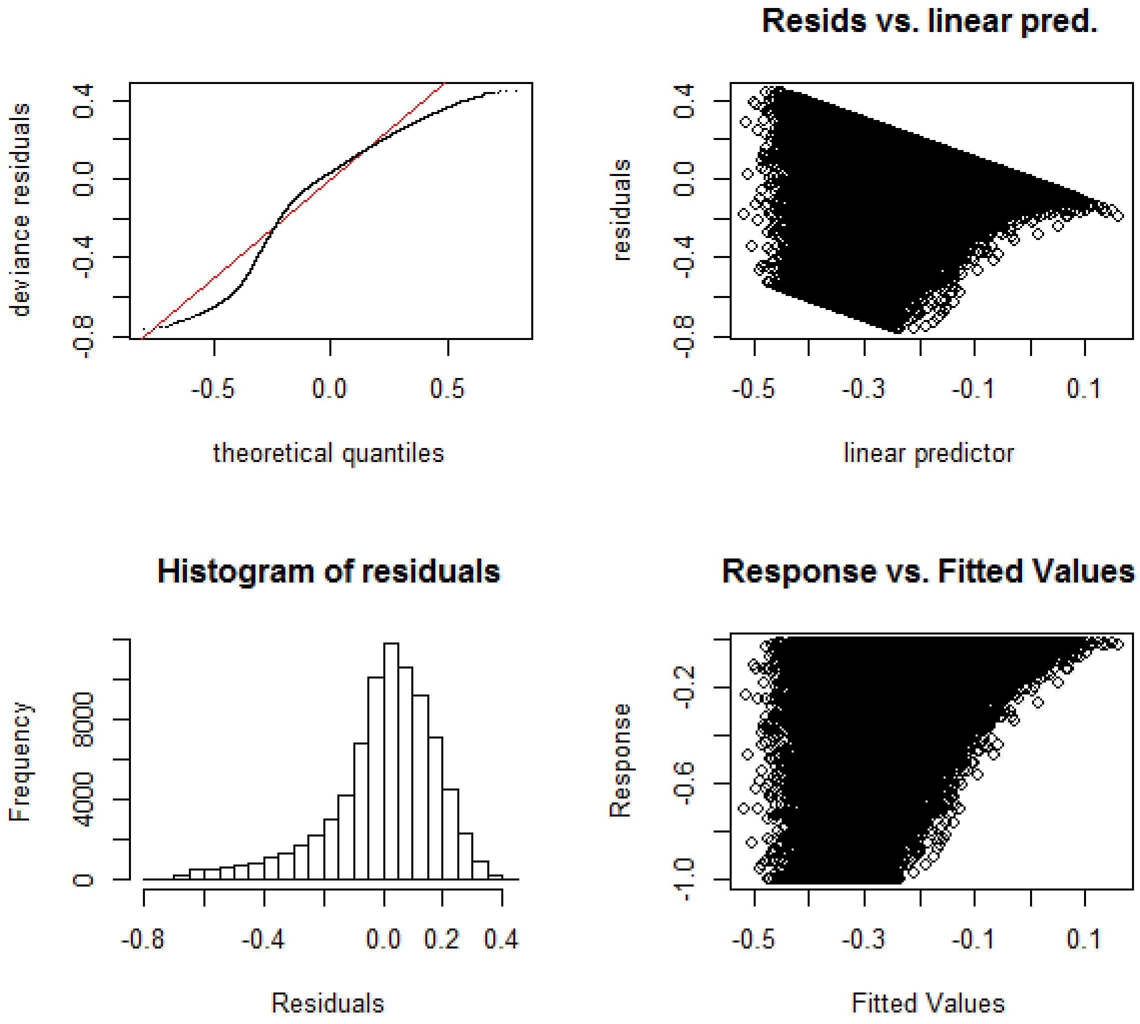

Figure 7.1 Checking plots for the full model.

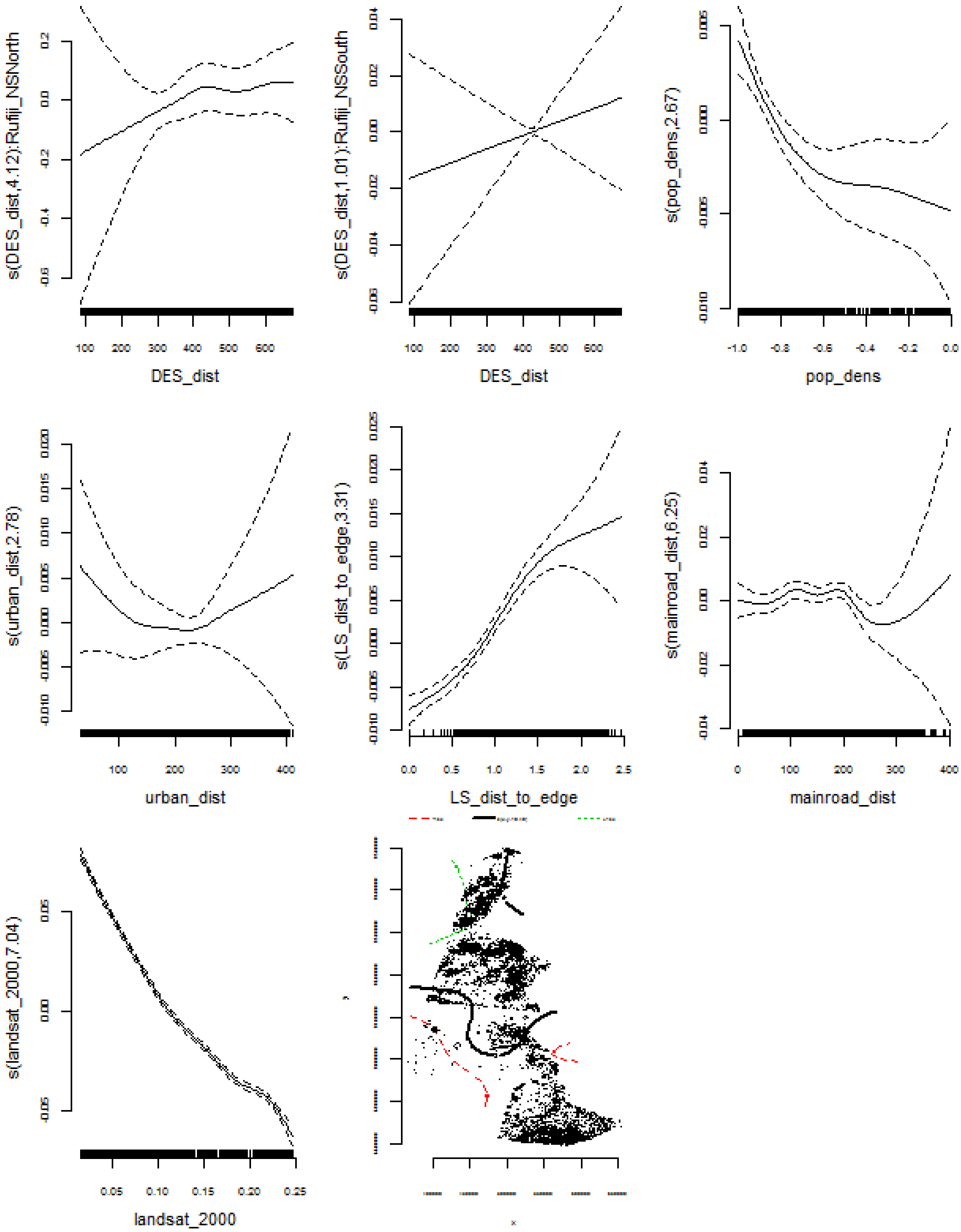

Figure 7.2 First quartile model (forest cover in 2000 < 0.25) 1 and 2 dimensional smooth functions against explanatory variables, with Bayesian confidence intervals shown as dotted lines.

Table 7.3 First quartile model summary

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-0.17399	0.058423	-2.978	0.00291	**
Rufiji_NSSouth	-0.04454	0.135376	-0.329	0.74217	
ethnoMakua	0.001973	0.003511	0.562	0.57404	
ethnoSwahili	-0.01188	0.009161	-1.297	0.19456	
ethnoWanyika	-0.01258	0.012334	-1.02	0.30797	
ethnoWashambala	-0.0186	0.009816	-1.895	0.05813	.
ethnoWayao	0.033871	0.014295	2.369	0.01783	*
ethnoWazaramo	-0.01249	0.009203	-1.357	0.17477	
regionsDaresSalaam	0.053856	0.045982	1.171	0.24152	
regionsLindi	0.054336	0.045064	1.206	0.22794	
regionsMorogoro	0.053201	0.046088	1.154	0.24838	
regionsMtwara	0.056237	0.044872	1.253	0.21012	
regionsPwani	0.05576	0.045875	1.215	0.2242	
regionsTanga	0.055727	0.046137	1.208	0.22712	
prot_areaYes	0.002319	0.001957	1.185	0.23613	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(DES_dist):Rufiji_NSNorth	4.119	4.404	0.395	0.829885	
s(DES_dist):Rufiji_NSSouth	1.01	1.014	0.553	0.458347	
s(pop_dens)	2.67	3.308	9.157	2.49E-06	***
s(urban_dist)	2.776	3.552	0.775	0.527328	
s(LS_dist_to_edge)	3.307	4.042	53.273	< 0.0000000000000002	***
s(mainroad_dist)	6.252	7.366	3.921	0.000221	***
s(landsat_2000)	7.037	8.078	1278.745	<0.0000000000000002	***
s(x,y)	19.48	23.091	3.486	3.01E-08	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.422 Deviance explained = 42.5%

fREML score = -24805 Scale est. = 0.001995 n = 14782

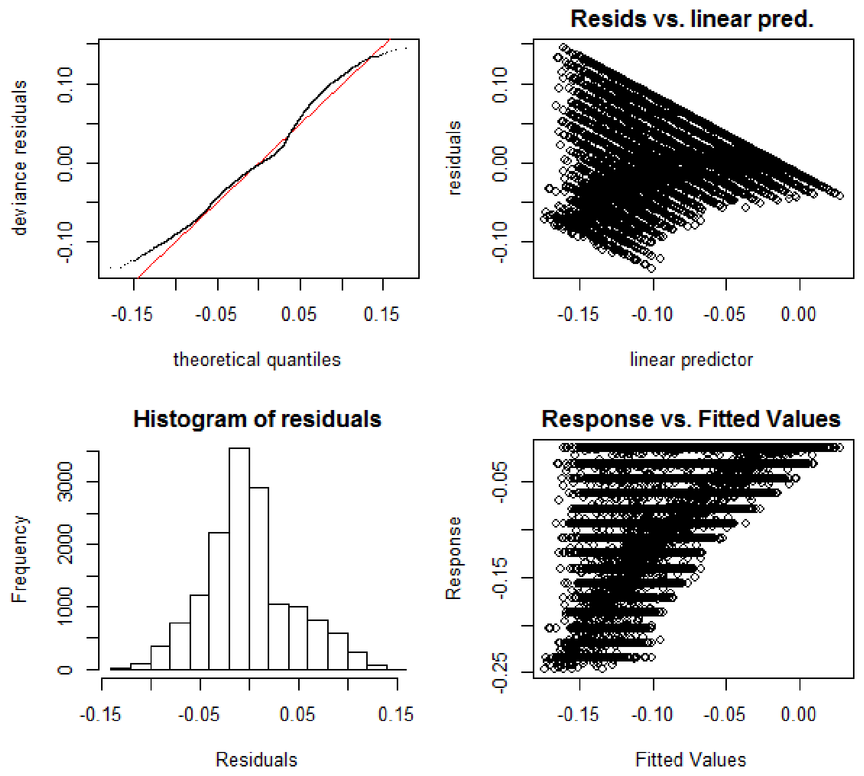

Figure 7.3 Checking plots for first quartile model (forest cover in 2000 < 0.25)

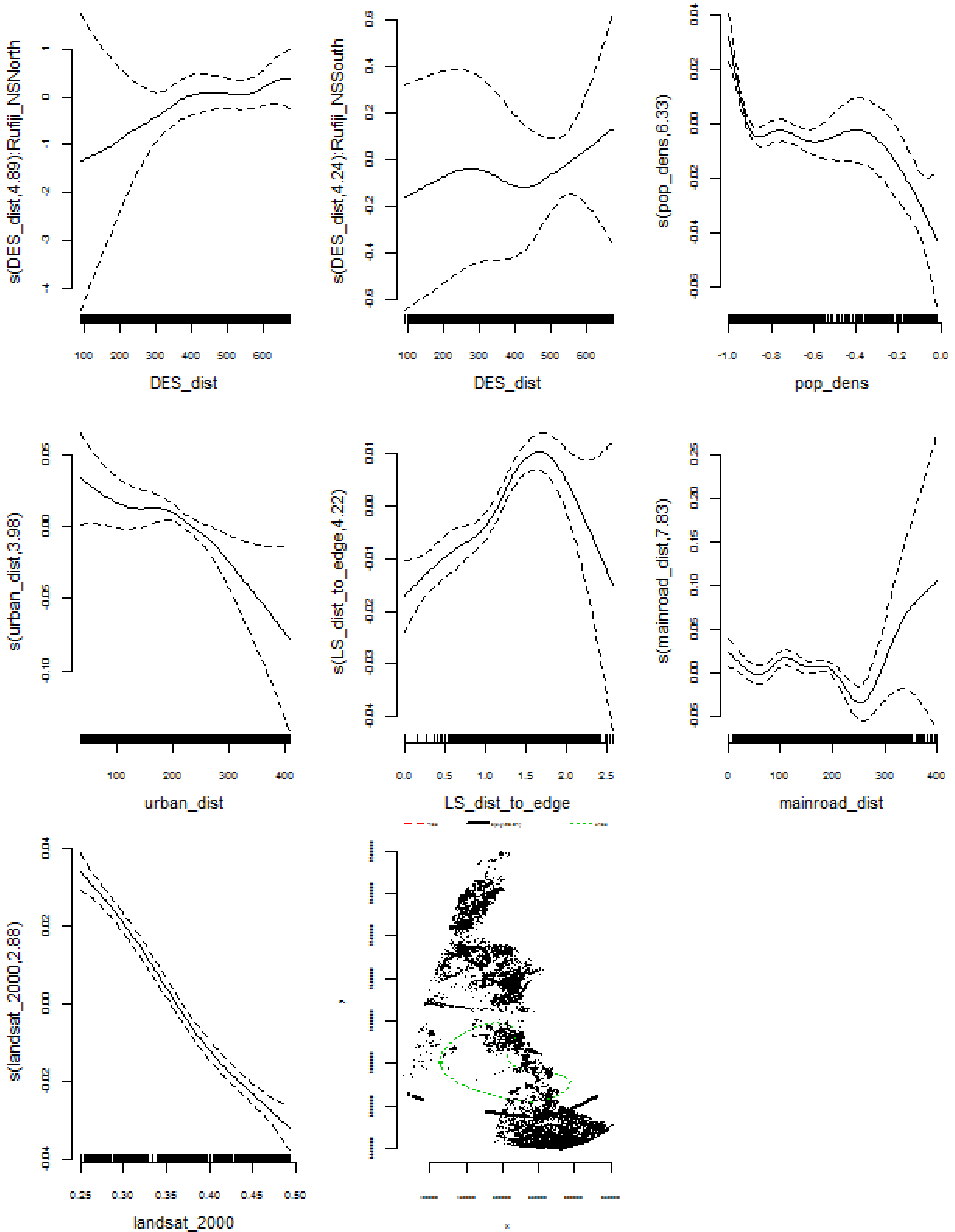

Figure 7.4 Second quartile model (forest cover in 2000 ≥0.25 and <0.5) 1 and 2 dimensional smooth functions against explanatory variables, with Bayesian confidence intervals shown as dotted lines.

Table 7.4 Second quartile model summary

Parametric	coefficients:				
	Estimate	Std. Erorr	t value	Pr(>|t|)	
(Intercept)	-0.693687	0.240572	-2.883	0.00394	**
Rufiji_NSSouth	-0.07035	0.96392	-0.073	0.94182	
ethnoMakua	-0.001191	0.008611	-0.138	0.89002	
ethnoSwahili	0.053871	0.029153	1.848	0.06464	.
ethnoWanyika	-0.053412	0.057957	-0.922	0.35676	
ethnoWashambala	0.039845	0.030798	1.294	0.19577	
ethnoWayao	0.076971	0.026974	2.854	0.00433	**
ethnoWazaramo	0.074773	0.02914	2.566	0.0103	*
regionsDaresSalaam	0.286103	0.113247	2.526	0.01153	*
regionsLindi	0.289169	0.110902	2.607	0.00913	**
regionsMorogoro	0.251448	0.113495	2.215	0.02674	*
regionsMtwara	0.2797	0.110597	2.529	0.01145	*
regionsPwani	0.304909	0.112436	2.712	0.0067	**
regionsTanga	0.364504	0.113172	3.221	0.00128	**
prot_areaYes	0.020638	0.004996	4.131	3.63E-05	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
Approximate significance of smooth terms:					
	edf	Ref.df	F	p-value	
s(DES_dist):Rufiji_NSNorth	4.89	5.317	1.148	0.3302	
s(DES_dist):Rufiji_NSSouth	4.241	4.957	2.147	0.0576	.
s(pop_dens)	6.327	7.412	11.33	6.80E-15	***
s(urban_dist)	3.981	5.118	1.875	0.0933	.
s(LS_dist_to_edge)	4.224	5.118	14.961	7.42E-15	***
s(mainroad_dist)	7.827	8.583	7.399	2.58E-10	***
s(landsat_2000)	2.881	3.58	150.11	2.00E-16	***
s(x,y)	23.813	25.802	7.314	2.00E-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
R-sq.(adj) = 0.422 Deviance explained = 12.5%					
fREML score = -24805 Scale est. = 0.001995 n = 14782					

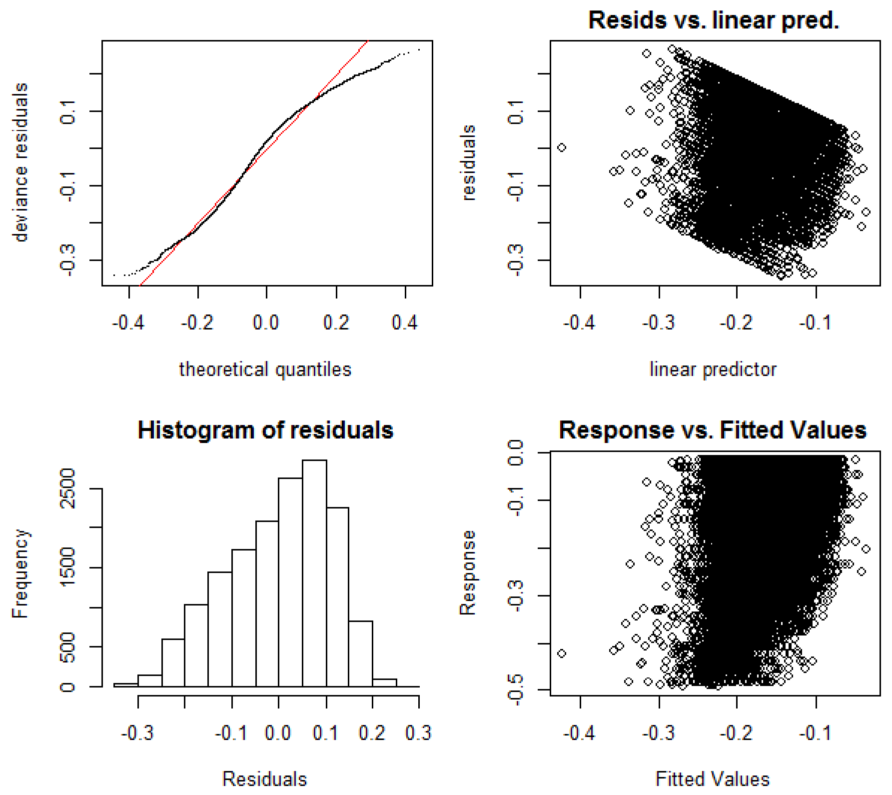

Figure 7.5 Second quartile model (forest cover in 2000 ≥ 0.25 and < 0.5) checking plots

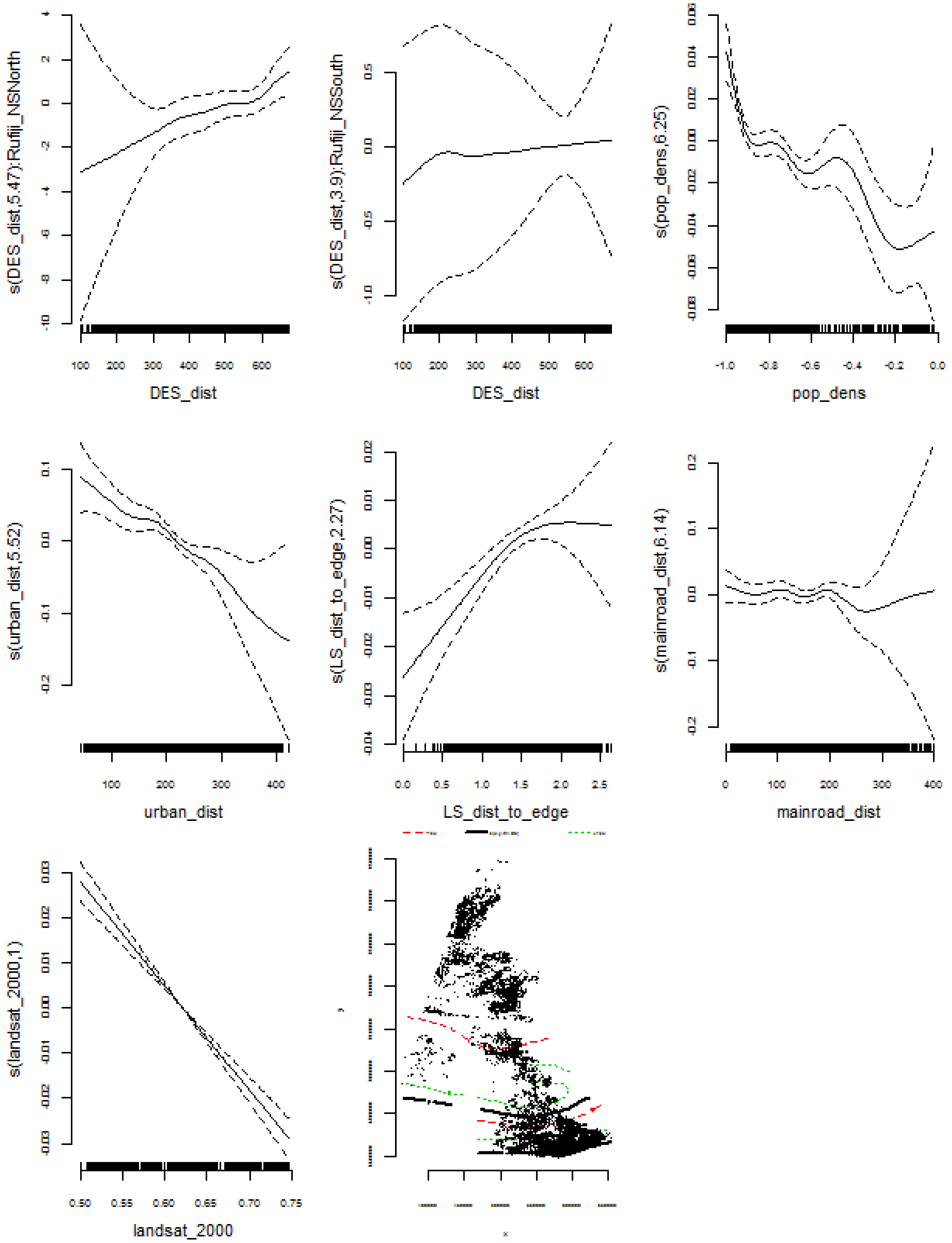

Figure 7.6 Third quartile model (forest cover in 2000 ≥ 0.5 and < 0.75) 1 and 2 dimensional smooth functions against explanatory variables, with Bayesian confidence intervals shown as dotted lines.

Parametric	coefficients:				
	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-0.82177	0.355328	-2.313	0.02075	*
Rufiji_NSSouth	-0.08116	2.203182	-0.037	0.97062	
ethnoMakua	-0.00253	0.015135	-0.167	0.86738	
ethnoSwahili	0.054284	0.045214	1.201	0.22992	
ethnoWanyika	-0.15577	0.117717	-1.323	0.18578	
ethnoWashambala	0.022326	0.047699	0.468	0.63975	
ethnoWayao	0.080174	0.03748	2.139	0.03244	*
ethnoWazaramo	0.113956	0.044216	2.577	0.00997	**
regionsLindi	0.04022	0.035441	1.135	0.25645	
regionsMorogoro	-0.08894	0.036747	-2.42	0.01551	*
regionsMtwara	0.014983	0.037095	0.404	0.68628	
regionsPwani	0.014387	0.025601	0.562	0.57415	
regionsTanga	0.059557	0.032079	1.857	0.06339	.
prot_areaYes	0.040043	0.007462	5.366	8.16E-08	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
Approximate significance of smooth terms:					
	edf	Ref.df	F	p-value	
s(DES_dist):Rufiji_NSNorth	5.467	5.803	4.5	0.000196	***
s(DES_dist):Rufiji_NSSouth	3.898	4.61	1.301	0.260456	
s(pop_dens)	6.248	7.32	13.168	2.00E-16	***
s(urban_dist)	5.518	6.809	4.611	4.92E-05	***
s(LS_dist_to_edge)	2.27	2.854	8.835	1.47E-05	***
s(mainroad_dist)	6.139	7.293	2.316	0.021686	*
s(landsat_2000)	1	1	163.675	2.00E-16	***
s(x,y)	24.933	26.613	11.151	2.00E-16	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1					
R-sq.(adj) = 0.422 Deviance explained = 10.2%					
fREML score = -24805 Scale est. = 0.001995 n = 14782					

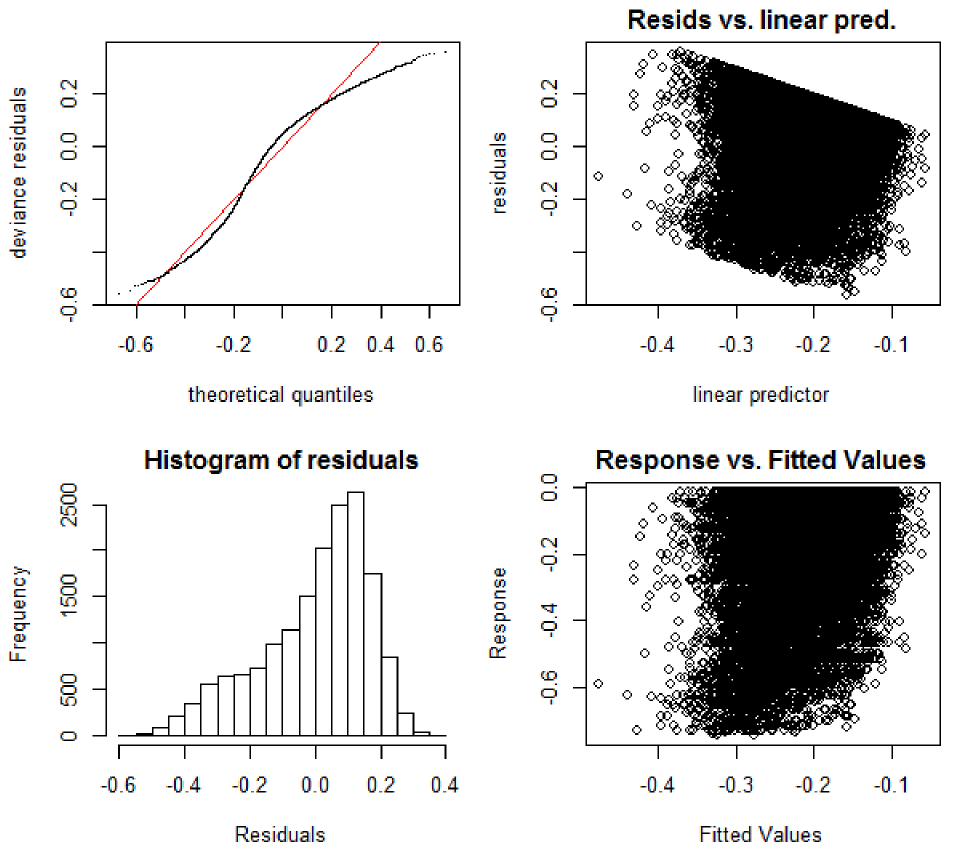

Figure 7.7 Third quartile model (forest cover in 2000 ≥ 0.5 and < 0.75) checking plots

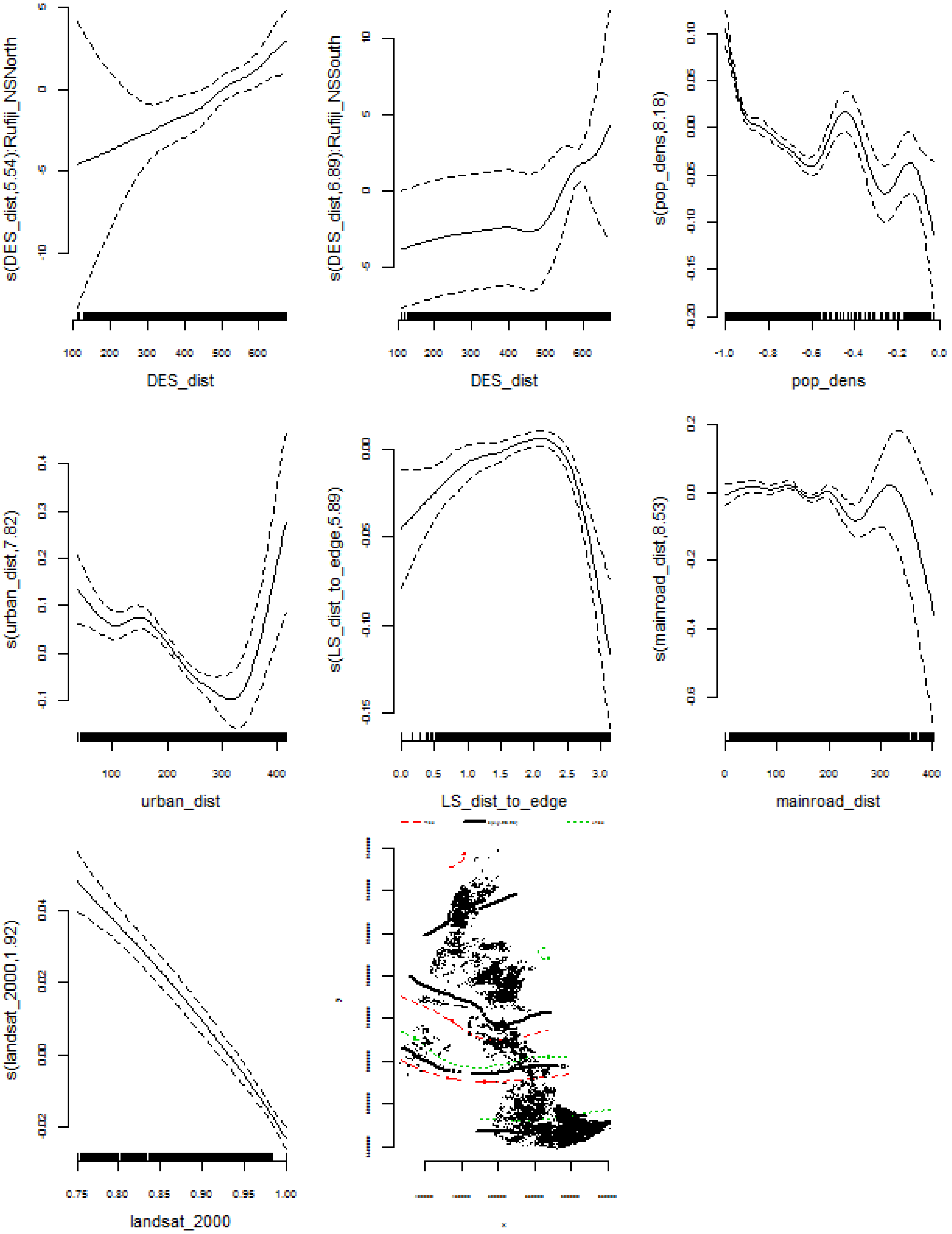

Figure 7.8 Fourth quartile model (forest cover in 2000 > 0.75) 1 and 2 dimensional smooth functions against explanatory variables, with Bayesian confidence intervals shown as dotted lines.

Parametric coefficients

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	2.942504	3.834026	0.767	0.4428	
Rufiji_NSSouth	-2.01278	3.748036	-0.537	0.5913	
ethnoMakua	-0.00295	0.02406	-0.123	0.9024	
ethnoSwahili	0.068006	0.057629	1.18	0.238	
ethnoWashambala	0.076437	0.061264	1.248	0.2122	
ethnoWayao	0.067293	0.037119	1.813	0.0699	.
ethnoWazaramo	0.12368	0.056955	2.172	0.0299	*
regionsDaresSalaam	-0.02256	0.245911	-0.092	0.9269	
regionsLindi	-0.0448	0.242253	-0.185	0.8533	
regionsMorogoro	-0.23732	0.246352	-0.963	0.3354	
regionsMtwara	-0.06004	0.242054	-0.248	0.8041	
regionsPwani	-0.10801	0.243732	-0.443	0.6577	
regionsTanga	-0.06943	0.244644	-0.284	0.7766	
prot_areaYes	0.051347	0.007297	7.037	2.00E-12	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(DES_dist):Rufiji_NSNorth	5.536	5.927	4.116	0.00043	***
s(DES_dist):Rufiji_NSSouth	6.893	7.064	5.589	1.75E-06	***
s(pop_dens)	8.179	8.764	37.552	2.00E-16	***
s(urban_dist)	7.82	8.608	10.053	8.62E-15	***
s(LS_dist_to_edge)	5.893	7.048	7.822	1.57E-09	***
s(mainroad_dist)	8.528	8.896	10.462	4.67E-16	***
s(landsat_2000)	1.918	2.367	109.206	2.00E-16	***
s(x,y)	26.251	27.703	18.943	2.00E-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.117 Deviance explained = 11.9%

fREML score = 159.97 Scale est. = 0.058327 n = 33742

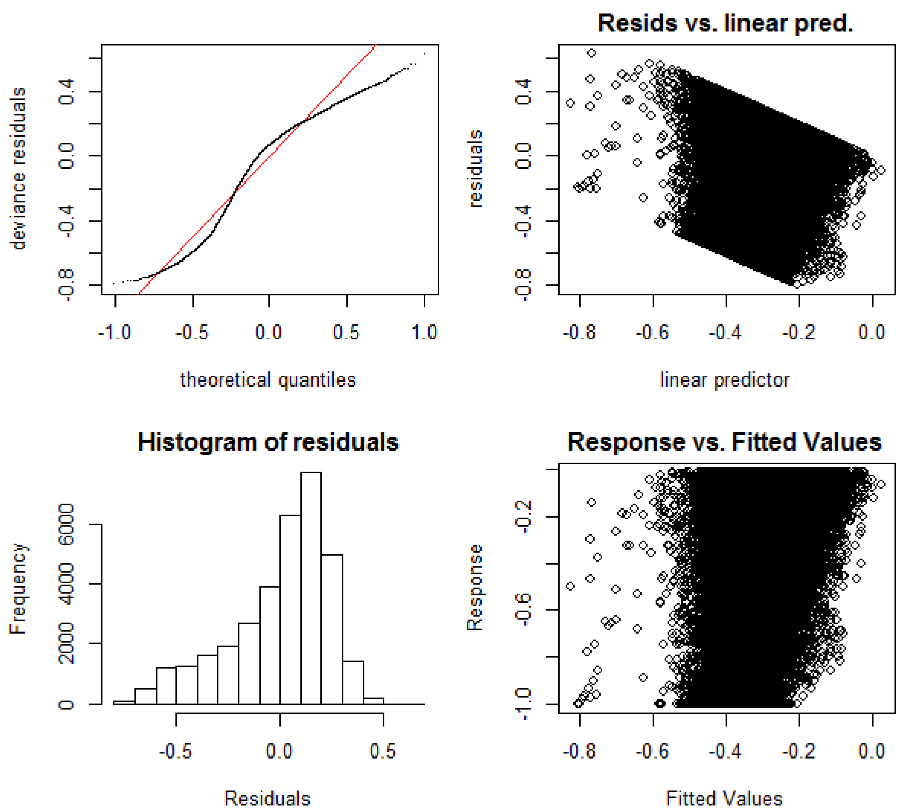

Figure 7.9 Fourth quartile model (forest cover in 2000 > 0.75) checking plots.