SPATIAL POINT PROCESS MODELLING OF LAND USE AND LAND COVER (LULC) CHANGE

Salma Anwar

PhD dissertation committee

Chair

Promoter

Prof. Dr. Ir. Alfred Stein	University of Twente

Assistant promoter

Dr. Ir. Wietske Bijker	University of Twente

Members

Prof. Dr. G. Câmara	INPE Brazil and University of Münster
Prof. Dr. E. Pebesma	University of Münster
Prof. Dr. Ir. M.G. Vosselman	University of Twente
Prof. Dr. V.G. Jetten	University of Twente

ITC dissertation number 257
ITC, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN: 978-90-365-3787-2
DOI: http://dx.doi.org/10.3990/1.9789036537872
Printed by: ITC Printing Department

© Salma Anwar, Enschede, The Netherlands
© Cover design by Benno Masselink
All rights reserved. No part of this publication may be reproduced without the prior written permission of the author.

UNIVERSITY OF TWENTE.

FACULTY OF GEO-INFORMATION SCIENCE AND EARTH OBSERVATION

SPATIAL POINT PROCESS MODELLING OF LAND USE AND LAND COVER (LULC) CHANGE

D I S S E R T A T I O N

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. H. Brinksma,
on account of the decision of the graduation committee,
to be publicly defended
on Thursday, November 06, 2014 at 14:45 hrs

by

Salma Anwar
born on April 01, 1977
in Lahore, Pakistan

This dissertation is approved by:

Prof. Dr. Ir. Alfred Stein (promoter)
Dr. Ir. Wietske Bijker (assistant promoter)

Summary

Land cover is defined as the biophysical state of the earth's surface and the immediate subsurface. Land cover changes are driven by natural forces or by land use by humans. Thus it involves both the natural and human dimensions. Locally, the land cover changes due to environmental or climatic factors determine the vulnerability of people to climatic perturbations and thus affect the decisions on land use by people. Similarly, the purpose for which a land is being used commonly has associated type of cover. Hence, understanding land cover changes is essential for understanding land use changes and vice versa. Most research on land use and land cover (LULC) change, however, has focused entirely on the human causes of LULC changes, ignoring the LULC changes due to natural factors. In this study I analyzed LULC changes due to natural forces and due to anthropogenic effects. I selected deforestation, being a major source of LULC change due to anthropogenic factors, for analyzing human influences on LULC change. As a comparison, I chose earthquakes for understanding the role of natural factors in LULC changes.

Disasters and land cover changes caused by earthquakes are initiated from point sources, i.e. locations of epicenters. Locations of earthquakes cluster on the globe and their distribution is determined by geological and geophysical characteristics of the epicenter region. Modelling the location pattern of earthquake epicenters requires a model that can best represent the clustering behavior of the earthquakes by taking into account the effects of the external factors. This study explored different spatial statistical techniques to analyze earthquake data in Pakistan recorded since 1973, including a major event in 2005. Modelling starts by identifying the Strauss point process model as the suitable model because of its flexibility to incorporate available geological information such as the presence of faults and plate boundaries as explanatory variables, and for its appropriateness to model this marked and clustered pattern. The results showed that the model is rigorous for its application to such a marked point pattern, representing well the clustering

behavior as determined by a number of environmental factors.

For deforestation studies, the historical patterns of forest LULC changes in the Brazilian Amazonia were analyzed. The large impact of LULC changes in this region on the world climate calls for an early and comprehensive assessment of its rate and pattern. To understand the current forest landscape structure, log landing sites (LLS) were detected from medium-resolution Landsat images. The spatial distribution of LLS served as a proxy to the intensity of selective logging activities. Selectively-logged forests were detected in the northern Rondonia state, north-western Mato Grosso state, and south-eastern Amazonas state in Brazil by applying spectral unmixing. End members used in this study represent basic components of a degraded forest environment. In the absence of field data, spectral unmixing of remote sensing images proved a suitable technique to detect forest degradation resulting from selective logging.

After detection of LLS from remote sensing images, the next study focused on the spatial assessment of the LLS patterns. The detected locations of LLS were represented by points and thus form a spatial point pattern. Spatial point pattern statistics served as an appropriate tool to study the LLS distribution. LLS data were of inhomogeneous structure and various ecological and geographic factors were considered to influence their distribution pattern. The objective was to discover important spatial characteristics of selective logging in the south-western part of the Brazilian Amazonia and then modelling the LLS pattern using a model suitable for inhomogeneous processes. A spatial model of LLS data should also include the interaction between the LLS in terms of their locations. The inhomogeneous J-function provided evidence of a strongly clustered LLS distribution. The Area-interaction model was capable to represent the clustered LLS pattern by incorporating the covariate information. The model showed a significant effect of presence of roads on the LLS distribution. The spatial statistical study proved helpful to quantify and better understand the LLS pattern.

To understand the temporal evolution of the dynamic LULC change, the proposed study was extended further to explore the potential of point pattern statistics in quantifying the spatio-temporal process of forest land cover change, as depicted from the remote sensing images, and determining a suitable model that could explain the behavior of forest LULC changes over time. For that purpose, we analyzed the spatial LLS patterns collected during the years 2000-2009 using spatial statistical methods. The purpose was to reveal important spatial and temporal characteristics of selective logging. After the spatial analysis, the patterns formed by the LLS were modelled using the higher-order Gibbs interaction models because of their suitability to model clustered patterns. The Area-interaction model and Geyer's saturation model

proved effective in modelling the clustered patterns in the absence of information about covariates. Results of both models conform closely to each other. We concluded that spatial statistical methods are powerful tools for analyzing and interpreting the spatial patterns formed by selective logging.

In all, spatial statistical methods proved very helpful for analyzing and modelling the LULC changes due to humans or due to natural factors. Point pattern statistics serves as a quantitative method which has a great deal of flexibility to analyze the initial patterns of LULC changes and is indispensable for analyzing and interpreting the LULC patterns in terms of a large number of ecological, geographic and socio-economic variables in an effective manner. Spatial statistics is, therefore, a promising way ahead for better understanding and then possibly reducing the adverse impacts of LULC changes on Earth.

Samenvatting

Landbedekking wordt gedefinieerd als de biofysische toestand van het aardoppervlak en de directe ondergrond. Veranderingen in landbedekking worden gedreven door natuurlijke krachten of door het landgebruik door de mens. Het gaat dus om zowel de natuurlijke als menselijke dimensies. Lokaal bepalen veranderingen in de landbedekking als gevolg van milieu- of klimatologische factoren de kwetsbaarheid van mensen voor klimatologische verstoringen en zijn dus van invloed op de beslissingen over landgebruik door mensen. Ook het doel waarvoor land wordt gebruik is vaak met een bepaald type van landbedekking geassocieerd. Vandaar dat het begrip van veranderingen in landbedekking essentieel is voor het begrijpen landgebruiksveranderingen en vice versa. Het meeste onderzoek op het gebied van veranderingen in het landgebruik en de landbedekking (Land Use and Land Cover, LULC), heeft zich echter gericht volledig op de menselijke oorzaken van LULC veranderingen, en negeert de LULC veranderingen als gevolg van natuurlijke factoren. In deze studie analyseerde ik LULC veranderingen als gevolg van natuurlijke krachten en door antropogene effecten. Ik koos ontbossing, zijnde een belangrijke bron van LULC veranderingen door antropogene factoren, voor het analyseren van de menselijke invloeden op LULC verandering. Ter vergelijking koos ik voor aardbevingen voor het begrijpen van de rol van natuurlijke factoren in LULC veranderingen.

Rampen en veranderingen in landbedekking veroorzaakt door aardbevingen worden geÃŕnitieerd uit puntbronnen, dat wil zeggen locaties van epicentra. Locaties van aardbevingen vormen clusters op de wereldbol en de verdeling ervan wordt bepaald door de geologische en geofysische kenmerken van het gebied rond het epicentrum. Het modelleren van het patroon in de locaties van de epicentra van aardbevingen vereist een model dat het best het clustering gedrag van de aardbevingen kan vertegenwoordigen door rekening te houden met de effecten van de externe factoren. Deze studie onderzocht verschillende ruimtelijke statistische technieken om gegevens van aardbevingen in

Pakistan te analyseren, geregistreerd sinds 1973, met inbegrip van een belangrijke beving in 2005. Modelleren begint met het identificeren van het Strauss punt-procesmodel als het geschikte model vanwege de flexibiliteit om beschikbare geologische gegevens te combineren, zoals de aanwezigheid van breuken en randen van (tektonische) platen als verklarende variabelen, en vanwege de geschiktheid om gemerkte en geclusterde patroon te modeleren. De resultaten toonden dat het een goed onderbouwd model is voor de toepassing op een dergelijk gemarkeerd punt patroon, en goed het clustering gedrag weergeeft zoals dat bepaald wordt door een aantal omgevingsfactoren.

Voor ontbossing studies, werden de historische patronen van bos LULC veranderingen in het Braziliaanse Amazonegebied geanalyseerd. De grote invloed van LULC veranderingen in deze regio op het wereld klimaat vraagt om een vroege en samenhangende beoordeling van hun snelheid en patroon. Om de huidige structuur van het boslandschap te begrijpen, werden veld depots (Loglanding sites, LLS) gedetecteerd op medium-resolutie Landsat beelden. De ruimtelijke verdeling van LLS diende als een proxy voor de intensiteit van selectieve houtkap.

Bossen met selectieve houtkap werden gedetecteerd in de noordelijke deelstaat RondÃť nia, noordwestelijke staat Mato Grosso, en zuidoostelijke staat Amazonas in BraziliÃń, door het toepassen van spectrale ontmenging. Eindelementen in deze studie vertegenwoordigen basiscomponenten van een gedegradeerde bosomgeving. In de afwezigheid van veldgegevens, bleek spectrale ontmenging van remote sensing beelden een geschikte techniek om de aantasting van bossen als gevolg van selectieve houtkap op te sporen.

Na detectie van LLS uit remote sensing beelden, richtte de volgende studie zich op de ruimtelijke beoordeling van de LLS patronen. De gedetecteerde locaties van LLS werden vertegenwoordigd door punten en vormden zo een ruimtelijk punt patroon. Statistiek van ruimtelijke patronen diende als een geschikt gereedschap om de LLS distributie bestuderen. LLS gegevens hadden een inhomogene structuur en de verschillende ecologische en geografische factoren werden verondersteld hun verspreidingspatroon te beÃŕnvloeden. Het doel was om belangrijke ruimtelijke kenmerken van selectieve houtkap in het zuidwestelijke deel van het Braziliaanse Amazonegebied te ontdekken en vervolgens de LLS patroon te modelleren met behulp van een model geschikt voor inhomogene processen. Een ruimtelijk model van LLS gegevens dient ook de interactie tussen de LLS te omvatten voor wat betreft hun locaties. De inhomogene J-functie leverde bewijs van een sterk geclusterde LLS verdeling. Het omgeving-interactie model was in staat om de geclusterde LLS patroon weer te geven door de integratie van de informatie over een co-variabele. Het model toonde een significant effect van de aanwezigheid van de wegen op de LLS distributie. De ruimtelijke statistische

studie hielp om het LLS patroon te kwantificeren en beter te begrijpen.

Om de ontwikkeling in de tijd van de dynamische LULC verandering te begrijpen, werd de voorgestelde studie verder uitgebreid om de mogelijkheden te onderzoeken van punt patroon statistiek bij het kwantificeren van de spatio-temporele proces van veranderingen in de bos landbedekking, zoals te zien is op de remote sensing beelden, en een geschikt model te bepalen dat het gedrag kan verklaren van veranderingen in bos LULC in de loop der tijd. Daartoe analyseerden we de ruimtelijke LLS patronen verzameld in de jaren 2000-2009 met behulp van ruimtelijke statistische methoden. Het doel was om belangrijke ruimtelijke en temporele kenmerken van selectieve houtkap te onthullen. Na de ruimtelijke analyse, werden de patronen gevormd door de LLS gemodelleerd met de hogere orde Gibbs interactie modellen vanwege hun geschiktheid om geclusterde patronen te modelleren. Het omgevingsinteractie model en GeyerâĂŹs verzadiging model bleken effectief in het modelleren van geclusterde patronen bij het ontbreken van informatie over co-variabelen. De resultaten van beide modellen stemmen goed overeen. We concludeerden dat de ruimtelijke statistische methoden krachtige gereedschappen zijn voor het analyseren en interpreteren van de ruimtelijke patronen gevormd door selectieve houtkap.

In het geheel bleken ruimtelijke statistische methoden zeer nuttig voor het analyseren en modelleren van LULC veranderingen veroorzaakt door mens of door natuurlijke factoren. Punt patroon statistiek dient als een kwantitatieve methode die een grote mate van flexibiliteit heeft om de oorspronkelijke patronen van LULC veranderingen te analyseren en is onmisbaar voor het effectief analyseren en interpreteren van de LULC patronen in termen van een groot aantal ecologische, geografische en sociaaleconomische variabelen. Ruimtelijke statistiek is derhalve een veelbelovende weg voorwaarts voor een beter begrip en vervolgens mogelijk beperking van de negatieve effecten van LULC veranderingen op Aarde.

Acknowledgments

Praise be to Allah, the most beneficent and the most merciful.

First of all I am very thankful to my country for providing us with funds to study in the Netherlands. I am grateful to Prof. Dr. Atta-ul-Rehman, the Ex-Chairman Pakistan Higher Education Commission, for initiating the overseas scholarships program. I am very thankful to the lovely country The Netherlands and its people for openly welcoming foreign students and providing us with ideal environment for studies and research.

To my first promotor Prof. Dr. Alfred Stein: I can not express enough my gratitude towards you. I am thankful to you for your kind supervision, thought provoking scientific suggestions and guidance. I learnt a lot from you in the fields of point pattern statistics, geo-statistics and circular statistics, apart from learning scientific attitude and scientific writing skills from you. I am thankful to you for always having responded to my e-mails very quickly and sending me quick feedbacks on work. You always gave me courage to overcome all kinds of troubles and motivated me to work whenever I started feeling like giving up or whenever my work was not progressing well. I have always felt amazed at how efficient, professional and dedicated you are with your work. I get inspirations from you and feel lucky to have worked with you.

I am also thankful to my second promotor Dr. Wietske Bijker for her technical guidance, thoughtful remarks, keen interest and encouragement during my research.

I am thankful to all my MSc teachers for teaching me the knowledge on subjects related to GIS and Remote sensing - the fields I had never studied before. I am thankful to you all for your moral support which I needed most during that period. I am thankful to Prof. John van Genderen for his cheering attitude and encouragement during my MSc thesis supervision.

I extend my deepest gratitude to all the people I met at INPE particularly Maria Isabel, Hilcea Ferreira, Thelma Krug and Dalton Valerionao

for their valuable discussions and support on my research. I am particularly thankful to Maria Isabel and Hilcea Ferreire for making my Brazil visit a memorable time of my life.

I feel greatly indebted to my friend and an excellent person Clarisse Kagoyire for always saying yes whenever I needed her for software related issues. I am thankful to Bas Retsios for helping me in programming. I feel highly obliged to Prof. Adrian Baddeley and Rolf Turner for their very kind response and always helping me whenever I wrote to them about sofware related problems.

My heartiest and sincere gratitude are paid to Theresa van der Boogaard, Marie Chantal, Bettine Geerdik and Marion Pierik for their professional as well as moral support. I am very thankful to Ginie Gerrits for always promptly responding to my travel related requests and making arrangements in almost no time. I am thankful to Helpdesk people, particularly Mr. Aiko Mulder, for their technical support. And thanks to Roeloff schoppers for his cheerful greetings every day. I am thankful to my office mates and friends Xudan Xu, Biao, Meisam and Hajer for their invaluable company, and to Sumbal, Coco, Liang and Rahul for their cheering talks. My special thanks are to my lovely friend Alka, her husband Amit and her daughter Bhuvan for having been like a blessing to me. I am also very thankful to Ms. Loes Minkman of NUFFIC for her appreciations and encouraging words which meant a lot to me.

Cordial thanks to all our Pakistani fellows at ITC particularly Amjad Ali, Haris A. Bhatti, M. Imran, M. Shafique, Mobushir R. Khan, Saleem Ullah, and Zahir Ali for their strong social support to us. Safiya Bhabhi, Sunbal, Amna, Nazneen and Kanza I miss the time we spent together during our all-ladies-parties. Thanks for those unforgettable times.

I am indebted most of all to my mother for all her sacrifices and contributions in my life. I am grateful to my father and my mother-in-law for their sincere prayers for our success. I am very thankful to Aasia for her big helping hand in taking care of my children particularly during my absence, and to Naveed and Siddique for always supporting us whenever and in whatever way we needed them. Thanks a lot to you! And to Zahida for showing how much you care about me and giving me tremendous support, even much more than I could ever expect from you, during last phases of my studies. God bless you and your family.

And finally to my husband Yaseen: you are an ideal life partner, and my children Asma, Taha and Sarah: I am very thankful to Allah for blessing me with you. Yaseen, Asma, Taha and Sarah you are my life.

Contents

List of Figures

List of Tables

General Introduction

1

1.1 Land cover and land use

'Land cover' and 'land use' are two important terms most often used to characterize the Earth's surface and the anthropogenic influence on it respectively. Land cover is defined as the biophysical state of the Earth's surface and the immediate subsurface (Turner and Skole, 1995). On the other hands, the term 'land use' refers to the purpose for which a land is used and hence is directly influenced by human activities and environmental processes and features (Silva et al., 2006). It involves both the manner in which the biophysical attributes of land are manipulated and the intent underlying that manipulation (Turner and Skole, 1995).

Land cover changes are driven by natural forces or by human land uses. Thus it involves both the natural and the human dimensions. Land cover affects land use. Locally, the land cover changes due to environmental or climatic factors determine the vulnerability of people to climatic perturbations and thus affect the decisions on land use by people. For example once the earthquake hazard maps are obtained, the area would be considered unsafe for human settlements and construction of sensitive establishments such as nuclear reactors etc. Globally, on the other hand, the land cover changes significantly affect the functioning of the Earth's system (Lambin and Geist, 2006). Similarly, the purpose for which a land is being used commonly has an associated type of cover e.g. agricultural, residential, or industrial. Therefore a change in land use also often causes changes in land cover. Hence an understanding of land cover change is essential for understanding land use change and vice versa. Since land use and land cover changes are closely interrelated and changes in land cover are mostly introduced due to changes in land use by humans, the terms of 'land use' and 'land cover' are also sometimes used interchangeably. In this study we will use the term "land use and land cover (LUCC)" for the combination of related use and cover.

Humankind has altered the terrestrial ecosystem to exploit the land

resources. Alterations on the Earth's surface bear strong implications to humans. Different land use activities like agriculture, cattle ranching, timber production, mining, leisure, habitation, and construction of transport infrastructure have the potential to bring about severe consequences to the environment and consequently to humans, if not planned and carried out efficiently. Therefore the global environmental agencies recognize the significance of LULC change studies. Researchers are working to find improved means for the measurement of LULC changes and understanding the factors behind the changes, for example, under the auspices of the Land-Use and Land-Cover Change (LUCC) project of the International Geosphere-Biosphere Programme (IGBP) and International Human Dimensions Programme on Global Environmental Change (IHDP) (Turner and Skole, 1995; Lambin, 1997). They seek to answer the questions such as: how has the land cover changed due to natural factors and what are the human contributions in changing the Earth's surface? How will changes in land cover and land use affect human life at the local, regional and global levels? How do the immediate biophysical and human dynamics affect the sustainability of land resources in different parts of the world? How does the land cover change affect the land use and vice versa? What are the driving forces behind the changes in land cover and land use and what role do they play in spatial and temporal dynamics of land use and land cover change? LULC studies are of interest to researchers, policy planners, and other decision makers because of implications of LULC change on global environment and sustainability, and utilization of land resources at local and global scales. Knowledge about LULC has become increasingly important as the world faces the need to overcome the problems resulting from deteriorating environmental quality.

While the LULC changes due to anthropogenic effects are well documented and several different methods and techniques have been developed to understand and model the dynamics of the land cover due to human influence, natural causes of land cover change have received much less attention. Data of LULC change caused by natural forces are much less analyzed, and their modelling has not been fully developed. Indeed, since the 1980's, most of the research on land cover change has focused entirely on the human causes of disasters. Nonetheless land cover changes occur in which the natural factors dominate. In such cases, the natural processes are responsible for all the damage whereas the human contribution is merely to have been in the way. There are circumstances in which human activity is also responsible for accelerating natural processes, but putting the blame entirely on human interference is a somewhat restricted point of view. LULC change studies should expand their mere understanding of social vulnerability to land cover changes by also considering the sensitivity and resilience

of the associated natural forces (Brookfield, 1999; Wisner et al., 2003).

In this study we chose deforestation for analyzing human influences on the process of LULC change, being a major source of LULC change due to anthropogenic factors. As a comparison, we chose earthquakes for understanding the role of natural factors in land cover changes.

1.2 Deforestation as a human cause of LULC change

The most striking example of land use resulting in land cover change is the clearing of large extents of forest lands for agriculture or cattle ranching. These changes carry huge consequences and bear strong implications to the local and global environment. Tropical rain forests particularly bear great importance because of their role in regulating the Earth's atmosphere and providing the regional population with a wide range of ecosystem goods and services. Since the last decades LULC changes in the tropics introduced by human activities have been driving strong environmental impacts by emission of greenhouse gases (Laurance et al., 1998), changes to forest biophysical properties (Koltunov et al., 2009), and loss to biodiversity and soil degradation (Upadhyay and Solberg, 2006). It also affects the lives of the people who depend directly on tropical resources for their economic and environmental needs (Vance and Iovanna, 2006).

Among all countries in the tropics, Brazil has the largest area of forest removed annually (Food and organization of the United Nations, 2010). The Brazilian rain forest covers an area of about 4 million square kilometers. The largest deforestation front is the well-known arc of deforestation in the Brazilian Amazon. Since the 1970's, approximately 16% of its original rain forest has been destroyed due to the government land ownership policies and a sophisticated road network running through the forest (Aguiar et al., 2007). The road network made access to the previously inaccessible interiors of the forest much easier to exploit for alternative land-uses such as cattle ranching, soybean production and selective logging for timber.

The accelerating LULC changes in the Brazilian forests raise serious concerns among international community due to its deteriorating effects on global climate, envisaging sincere efforts to curb the current rate of LULC change. The enormous potential impact of LULC change in the Amazonia calls for an early and comprehensive assessment of the magnitude, spatial scale and pace of LULC change, as well as the socio-economic factors responsible for it.

Deforestation as a LULC change is not an event but a process (INPE, 2011b). Conversion of primary forest to the stage of clear-cut deforestation may take between several months and several years depending

upon the activity influencing the forest structure. Each particular activity shapes the forest land cover in a specific way which can be detected from remote sensing data and can be related to a specific land use. The spatial patterns of forest land cover thus can serve as important indicators of land use activities and their driving factors (Lorena and Lambin, 2009).

In the Amazon, two main sources of LULC change occur: deforestation by forest degradation and deforestation by clear cutting the patches of land. They can be distinguished by the length of period between the times of occurrence and the opportunity for its detection as LULC change. The LULC change introduced by forest degradation is a gradual process which can be observed as a continuous and partial loss of forest cover characterized by a change in the forest canopy structure and its spectral reflectance in remote sensing imagery. Deforestation as distinguished by clear cutting, however, is the process of forest cover removal in a short period of time. This process is visible in remote sensing images, characterized by the predominance of exposed soil with a smooth texture. Deforestation resulting from clear cutting of the forest for the purpose of cattle ranching or agricultural production has received much attention worldwide as compared to deforestation by forest degradation.

Two types of forest degradation can be distinguished: degradation caused by natural events or degradation caused by anthropogenic effects. Natural causes of forest degradation include natural fires, droughts and hurricanes. We will not deal with the natural causes because they cannot be easily influenced by policy interventions, and also their occurrence and contribution to the net forest degradation, and hence deforestation, is much limited as compared to that due to the human factors (Contreras-Hermosilla, 2000). Anthropogenic effects mainly consist of selective logging activities. Interactions between loggers and follow-on farmers which results in widespread deforestation, has long been reported (Walker, 1987; Lambin, 1997; Rudel and Roper, 2002). Selective logging has deteriorating effects on the forest environment including changes in the light regime and forest micro-climate, erosion, soil compactness, disruption of nutrient cycling and possibly long term changes in tree species composition (Koltunov et al., 2009). Selective logging also affects the diversity of forest fauna and increases the susceptibility of forest to fire through modification of the under-canopy micro-climate and supply of fuel in the form of debris left after logging (Gerwing, 2002). It is estimated that timber extraction and logging activities account for more than 70% of total forest degradation in Latin America and tropical Asia (kissinger et al., 2012). Moreover in a study on effects of selective logging in the Brazilian Amazon, selectively logged areas were found to be four times more likely than unlogged areas to be fully deforested in

subsequent years (Asner et al., 2005).

While whole patches of land cleared for cattle ranching or agricultural use are observable from satellite imagery, selective logging is hard to monitor as it causes a spatially diffuse thinning of forest canopy structure and it needs high resolution data and special image mining techniques for its detection (Asner et al., 2005). Spatially explicit diagnostic models of LULC change can then be developed from the remotely sensed observations of LULC patterns (Mertens and Lambin, 1997). However, to understand the driving forces of such observed changes, the remotely sensed data must be linked to socioeconomic data. The LULC change models that link remote-sensing observations with ground-based social data can greatly improve our understanding of the determinants of LULC change. Development of spatially explicit LULC change models based on the remote-sensing data in combination with environmental and socio-economic factors is also required for understanding the cause-effect relationship between driving forces and LULC change.

1.3 Study area 1

The study area (Fig. 1.1) selected for this study is located in the south-western Brazilian Amazonia, covering 30,000 km^2. It constitutes the northern Rondonia state, north-western Mato Grosso state, and south-eastern Amazonas state – three of the top five timber-producing states (Asner et al., 2005) which account for about 95 percent of the region's deforestation (INPE, 2011a).

The region is drained by the Machado river, which a tributary of the Madeira river (the biggest tributary of the Amazon). The terrain is undulated, ranging from 100 to 450 m above sea level. The climate in the area is classified as equatorial hot and humid, with tropical transition. It has a dry season that lasts from June to August. The annual average precipitation is 2016 mm and annual average temperature is 25.5^{o}C (Lu et al., 2004).

The area was part of a colonisation project initiated by the Brazilian government in the 1970's. The project has played a major role in deforestation of the area. The project aimed at providing land to the landless people from the south-eastern region of Brazil. Migrants from the south converted forested areas into crop lands. The region has, therefore, been deforested at high rates.

The study area has been attractive for timber loggers since long. It lies in the transition forest zone where commercially valuable timber trees are found in lower density (app. 20 m^3 ha^-1) than in the dense forests (35-40 m^3 ha^{-1}) (Monteiro et al., 2003). The loggers build roads in the area for carrying out mechanized logging operations (Lu et al.,

2004). Along the roads, the logged timber is stored temporarily from where it is gathered and exported to the markets (Stone and P., 1998; Souza and Barreto, 2000; Matricardi et al., 2005). These log-landing sites (LLS) may serve as spatial signature of areas under selective logging. The LLS and their distribution patterns can be useful in determining the processes operating in the study area related to the logging activities.

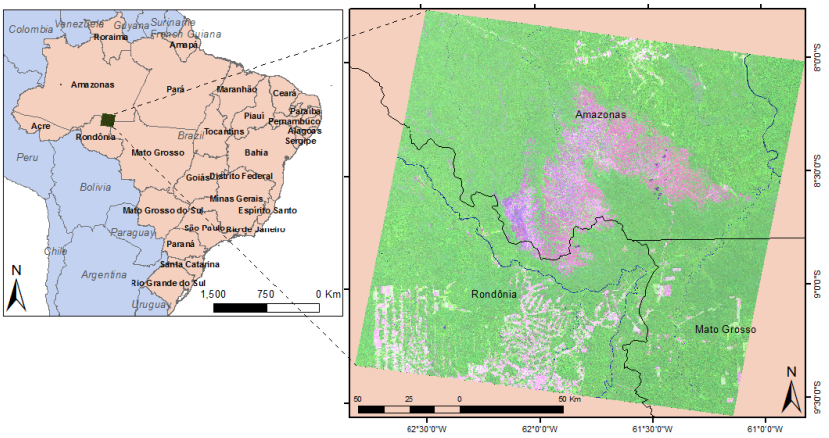


Figure 1.1: Location of the study area in the north-west of Brazil and the Landsat-5 TM image corresponding the area.

1.4 Earthquakes as natural cause of LULC change

Earthquakes represent one of the most devastating and destructing natural forces altering the earth's surface. LULC changes, induced by large physical disturbances such as earthquakes, can influence human life in adverse ways. Earthquakes affect terrestrial ecosystems and trigger landslides and other geological movements which continue to occur for some time in the earthquake-affected area. Compared to volcanic eruptions, hurricanes and other nature disasters, earthquakes may have a greater potential to significantly change landscape patterns in the affected areas, particularly in the area near the epicenters. An earthquake of a high magnitude can induce a huge number of geological hazards withsin the scope of 100,000 km^2 from the epicenter location.

Earthquakes have a dramatic impact on the lives and well-being of people at many places on the globe. Recent and continued population growth, development and the expansion of life over hazardous areas are increasing the potential impact of earthquake disasters. Due to their

vast destructions, there has been an ongoing urge to understand the earthquake dynamics and identify causes of earthquake occurrences so as to save mankind from their deleterious effects. Obtaining this knowledge is crucial to assess the tectonic processes and their potential impacts. Ultimately earthquake hazard maps can be obtained to define precautionary measures to mitigate the destructiveness of future earthquakes. Such information is also of great economic importance, on regional as well as on global scales, since recommendations for infrastructure construction are based on such studies. Land uses in the earthquake prone areas such as the settlement of major cities and construction of critical facilities, e.g. nuclear reactors, highlight the importance of securing the human populations against the earthquake hazards, disaster mitigation and prevention (Vere-Jones, 1995). Land cover changes resulting from earthquakes are therefore critical for the study and detailed understanding of tectonics in the affected area. The analysis of epicenter locations plays the foremost role in analyzing the seismicity of an area. Distributional characteristics of epicenter locations can be very helpful in understanding the distinctive geological and geophysical characteristics of the seismic locations which in turn can give insight into the mechanism responsible for the earthquake distribution pattern (Zheng and Vere-Jones, 1991).

1.5 Study area 2

For analyzing and modelling the distribution pattern of earthquake epicenter locations and investigating the role of available geographical and geophysical factors in determining the locations of earthquake occurrences, we chose seismicity record of the Northern region of Pakistan (Fig. 1.2). The region is an active seismic zone. It lies on the western edge of the Himalayan Arc, which denotes the region of continental convergence between the Indian and Eurasian tectonic plates. Location of tectonic plate boundary plays a significant role in determining seismicity of the study area. Compression motion between the two tectonic plates results in a series of large thrust faults in the study area including the Main Karakorum Thrust (MKT), the Main Boundary Thrust (MBT), and the Main Mantle Thrust (MMT) (Naranjo, 2008).

The seismic record of the area was taken from the USGS website (USGS, 2008). The data included 1403 earthquakes that occurred in the region between January 1973 and August 2008. The temporal distribution of the data showed that a larger seismic activity in the region occurred in 2005 than in other years. This is due to the Kashmir earthquake of magnitude 7.6, which struck the region on October 8, 2005, and was followed by a series of aftershocks. Most of the aftershocks of

Kashmir Earthquakes were concentrated around the north-western tip of the earthquake causative fault (Ömer Aydan, 2006).

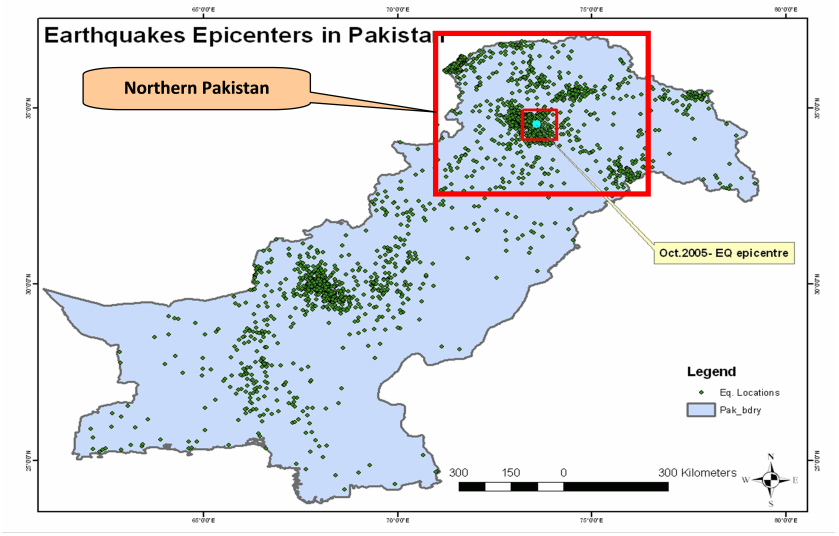


Figure 1.2: Distribution of earthquakes (represented by points) on Pakistan map. The red rectangle constitutes 1403 earthquakes recorded between January 1973 - August 2008 in the northern Pakistan

1.6 Modelling LULC change

LULC change models are needed for the analysis of environmental processes and problems that must be understood if living conditions and standards are to be improved or maintained at current levels. A wide range of land-use change models, for different scales and research questions and based on a variety of approaches is available (Brown et al., 2004). Different models of LULC change address different questions, for example, location of change versus quantity of change; some models consider an area of land as the unit of analysis, while others are centered on individuals as decision making agents. LULC change modelling requires integrated global and regional modelling approaches informed by empirical assessments of the patterns of LULC change to help us improve our understanding of how the processes of LULC change vary across spatial and temporal scales. Spatial models have not been widely developed and used, in spite of the necessity of spatial data for answering LULC change related ecological questions. Spatially explicit

modelling is an important technique to help us understand of the key LULC change processes, and for describing them in quantitative terms. The lack of development of spatial models is probably because the data and computational requirements have, in the past, been prohibitive.

For effective LULC change monitoring, it is of vital importance to take historical patterns into consideration which can ultimately help in understanding the current LULC change structure. Historical LULC change data allow researchers to study past relationships between landscape change and ecological processes. Initial configuration of landscape elements is one of the basic three components of all models of landscape change as identified by (Baker, 1989). The locations of initial LULC may point to the factors that are most responsible for the expansion of LULC. Analysis of historical patterns of LULC then starts from the locations where for example a few trees were initially cut for the purpose of logging. Study of these locations is important as with the passage of time they grow in size leading to a widespread deforestation. Due to significant contribution of selective logging in worldwide deforestation rates, forest degradation following selective logging has been termed "cryptic deforestation" by nepstad et al. (2005). The deforestation literature, for example Walker (1987); Rudel and Roper (2002), have long recognized interactions between loggers and follow-on farmers for their contribution in deforestation process.

Although logging causes less abrupt changes to forest structure as compared to clear- cutting, selective logging causes forest fragmentation and makes the logged areas more vulnerable to forest fires and soil erosion and ultimately deforestation. In the analysis of land cover changes, such land cover modifications i.e. subtle changes that affect the character of the land cover without changing its overall classification are as important as land cover conversions i.e. the replacement of one cover type by another (Lambin and Geist, 2006). There has been intensive work on analyzing the deforestation patterns and processes in the Brazilian Amazonia and efforts have been made to model the deforestation processes using different spatial analytical tools. Spatially explicit analysis and modelling of the locations of forest degradation is, however, lacking.

In the study of LULC changes resulting from earthquakes, the earthquake epicenter locations, like the selectively logged sites, bear great significance in analyzing the factors responsible for earthquake occurrence and demarcating the earthquake prone regions. The disasters caused by earthquakes are also initiated from point sources, i.e. locations of epicenters. Locations of earthquakes cluster on the globe and their distribution is determined by geological and geophysical characteristics of the epicenter region. Modelling the data constituting the locations of earthquake epicenters requires a model that can best rep-

resent the clustering behavior of the earthquakes by take into account the effects of the external factors.

Modelling earthquake data has since long been a focus of research by seismologists and statisticians (Utsu, 1961; Kagan and Knopoff, 1977; Vere-Jones and Li, 1997; Ogata, 1998; Zhuang, 2000; Schoenberg, 2003) to predict the behavior of the earthquakes and their aftershocks occurrences so as to be able to safeguard humanity against the vast destruction and panic caused by earthquakes. The current analysis aims at finding and analyzing the prominent features of the available data set for the Kashmir earthquake 2005 and its aftershocks, and analyzing and modelling the influence of the available geological and geophysical variables in determining the distribution pattern of the earthquake aftershocks.

Due to the point-like spatial locations of the initial stages given by epicenters and selective logging, the spatial point processes serve as a relevant and suitable tool for investigating and modelling distributional structure of early land cover changes caused by earthquakes and selective logging, respectively. The locations of earthquake epicenters and selective logging sites serve as good examples of spatial point pattern since the these locations can be regarded as two dimensional points (X, Y) in geographical space. The pattern is presumed to be formed as a result of some form of stochastic mechanism (spatial point process). A point process and a point pattern are two different concepts in the sense that the point process is a stochastic model and the point pattern is a realization of the process (Isham, 1984; Perry et al., 2006). Various researchers have put extensive contributions in defining different types of point processes and providing theoretical understandings of the processes (Cox and Isham, 1980; Baddeley and Silverman, 1984; Högmander and Särkkä, 1999; Daley and Vere-Jones, 2002; Diggle, 2003; Møller and Waagepetersen, 2007; Vere-Jones, 2009).

The chosen point patterns may represent the two different classes of point processes described by Cressie (1993). The earthquakes example illustrates one case, in which successive events are treated as individual points in both space and time. Whereas the LLS distribution can be thought of as a spatial point pattern which forms a trajectory in space-time. An effort to model LULC patterns of both types requires a basic understanding and incorporation of the physical processes underlying the initiation and subsequent evolution of the LULC events in a spatially explicit manner. Although the physical processes responsible for the two selected phenomena are quite different, the two data sets have many features in common which make the spatial statistical techniques chosen for modelling and addressing the physical processes equally and adequately applicable to both cases. For the LLS distribution, however, acknowledging the temporal evolution of the process is also important

for a correct and insightful interpretation of the current and future LULC dynamics. Therefore for the LLS data, we performed a spatial-temporal analysis by detecting the changes in spatial patterns of LULC change over a period of 2000-2009.

1.7 Research objectives

Analyses and modelling of land use and land cover (LULC) changes due to human causes and due to natural forces using spatial point pattern statistics.

This main objective is divided into sub-objectives as below:

1. To analyze and model the spatial distribution of earthquake epicenters as natural cause of LULC change
2. To detect the spatial distribution of selectively logged sites from remote sensing images
3. To analyze and model the spatial distribution of selectively logged sites as human cause of LULC change using spatial point pattern statistics
4. To analyze spatial-temporal behavior of the distribution of selectively logged sites using spatial point pattern statistics

1.8 Thesis structure

The dissertation comprises 6 chapters including 4 technical chapters, out of which three are published in or submitted to peer-reviewed ISI journals and one is published as a book chapter. In general format, each chapter is arranged as abstract, introduction, methodology, results, discussion and conclusion sections.

Chapter 1 gives a detailed overview and scope of the dissertation. It includes general introduction, motivation for the research work, research objectives, relevance of the research and thesis roadmap.

Chapter 2 presents different spatial statistical techniques to analyze earthquake data in Northern Pakistan recorded since 1973, including a major event in 2005. For earthquakes, modelling starts by identifying a suitable model. In particular, the Strauss point process model has been investigated for its exibility to incorporate available geological information such as the presence of faults and plate boundaries as explanatory variables, and for its appropriateness to model the marked and clustered pattern of earthquake aftershocks.

Chapter 3 demonstrates the effectiveness of spectral linear unmixing technique for detection of selective logging in a Landsat 5 TM image constituting parts of northern RondÃť nia state, north-western Mato Grosso state and south-eastern Amazonas state in the Brazilian Amazonia. Selectively logged forests were detected by applying spectral unmixing. Afterwards, spatial statistical analysis of the detected locations was performed to detect important features of the distribution of selectively logged locations in the study area.

Chapter 4 demonstrates the need and usefulness of the modified J-summary function to analyze the non-homogeneous structure of a point pattern consisting of selectively logged locations in the study area. Choice of a suitable kernel bandwidth in modelling the non-stationarity was emphasized. After investigating the inhomogeneity of the logging pattern, the Area-interaction point process model incorporating information about distance of logging sites to roads and to clear-cut deforested areas was applied to explain the distribution of selective logging and contribution of the explanatory factors found significant at the scale of the area covered by a Landsat scene.

Chapter 5 presents the analysis of the spatial patterns of the selectively logged sites collected during the years 2000-2009 in a part of the Brazilian Amazonia, using spatial statistical methods. The purpose was to reveal important spatial and temporal characteristics of selective logging. After the spatial analysis, the spatial patterns formed by selective logging were modelled using the higher-order Gibbs interaction models due to their suitability to model clustered patterns.

Chapter 6 presents a synthesis of the results produced during the research work, and their discussion. It also provides the main conclusion of the research and its significant contribution in scientific knowledge. Finally, future research directions and recommendations are provided.

The Strauss point process model for earthquake aftershocks

2

Abstract

Environmental spatial point processes such as earthquake epicenters are usually related to a range of ambient factors. Epicenters can be seen as marked point processes with respect to their location and size. Large data sets exist that include recordings with covariates that may possibly influence their occurrence, magnitude and location. This study explores different spatial statistical techniques to analyze the pattern of earthquake aftershock epicenters in Pakistan recorded since 1973, including a major event in 2005. Spatial modelling of the epicenters starts by identifying a suitable model. In particular, the Strauss point process model has been investigated for its flexibility to incorporate available geological information such as the presence of faults and plate boundaries as explanatory variables, and for its appropriateness to model this marked and clustered pattern. The results show that the model, despite some limitations, is rigorous for applying it to such a marked point pattern, representing well the clustering behaviour as determined by a number of environmental factors.

2.1 Introduction

Earthquakes are disasters that apparently occur at erratic seismic locations and at unexpected moments. They may cause vast destruction and panic among the affected population. Processes generating earthquakes are prominent in earthquake prone areas that are least partly

[1]This chapter is based on the following paper

Anwar, S., Stein, A., van Genderen, J. L., 2012. Implementation of the Marked Strauss point process model to the epicenters of earthquake aftershocks *In: Shi W, Goodchild M, Lees B, Leung Y (eds) Advances in geo-spatial information science. Taylor Francis, London 13 (3), 125-140.*

determined by geological faults and occur in particular close to subduction zones. Modelling earthquake data has since long been a focus of research by seismologists and statisticians. A spatial point pattern analysis of earthquake epicenters can be helpful to improve our understanding of location specific geological and geophysical characteristics (Zheng and Vere-Jones, 1991). Having a better knowledge on where the earthquakes, e.g. as major events or as aftershocks occur in relation to geological features, may thus result in identification of hazard zones (Li and Vere-Jones, 1997; Zhuang, 2000; Schoenberg, 2003).

A statistical point pattern analysis usually consists of three steps. The first step is to estimate the intensity function followed by modelling the dependence of the points representing the process. Modelling of the spatial dependence is usually done by applying the G-function. The second step concerns identification of a suitable model that explains the relation between the locations and geological features (Albert et al., 2002; Turner, 2009a). Modelling and extracting information from a point pattern with multiple types of marks, also suspected of being influenced by covariate effects remains a challenge in point process statistics (Illian et al., 2008). Various researchers have put extensive contributions in defining different types of point processes and providing rich theoretical understandings of the processes (Cox and Isham, 1980; Cressie, 1993; Högmander and Särkkä, 1999; Daley and Vere-Jones, 2002; Diggle, 2003; Baddeley et al., 2005; Illian et al., 2008; Gelfand et al., 2010). The third step concerns the validation of the model, including model diagnostics.

For epicenters, the Poisson point process model is used in the absence of dependence among epicentres. It may investigate the effects of covariates using likelihood ratio tests (Diggle, 2003). If dependence or interaction exists among epicenters, however, the Poisson model falls short and the likelihood ratio tests may fail to fit realistic model to the data due to its inability to determine the normalizing constant found in the intensity function of such processes. In those circumstances, the Strauss point process model may serve as an alternative, as it is sufficiently rich to describe the interaction among epicenters, whereas the Poisson process model may serve as a reference to test the data for the presence of interaction. The Strauss point process (Strauss, 1975; Kelly and Ripley, 1976) is an example of pairwise interaction models that can deal with modelling inter-epicenter interactions. Model diagnostics include different types of formal and informal techniques found in literature.

The aim of this study is to explore the possibilities for carrying out a spatial statistical study on the aftershock epicenters following the major Kasmir earthquake that occurred in Pakistan, in 2005. The analysis aims at finding and analyzing the epicenters as a marked point pattern. It explores the use of the Strauss point process model to represent the

distribution of epicenters in a well-defined geographical area.

2.2 Statistical methodology

2.2.1 Directional Statistics

Directional statistics relate the occurrence of events towards one central point (Fisher, 1996; Mardia and Jupp, 1999). A division is made between circular and spherical data. In this study we concentrate on treating epicenters as circular data, since the depths of almost all the earthquakes hypocentres were recorded to be constant as 10km below the earth's surface. Essential is the distribution of the angles of the directions between successive epicenters. A typical feature of directional data is that the jump from 360 to 0 degrees is carefully dealt with when fitting spatial distributions. Common tools in directional statistics are the rose diagrams and the distributions like the von-Mises distribution that allow appropriate fitting to the data (Fisher, 1996).

2.2.2 Intensity and the $G-$function

Epicenters form a spatial point pattern. Point patterns are the result of mixture of first order and second order effects. First order effects are described by the intensity. The intensity is a localized first moment measure given by the expected number of points per unit area. Intensity may be constant (homogeneous) or spatially varying (inhomogeneous) throughout a study region. Second order effects are given by inter-point interactions which describe the stochastic dependence between points. Thus mathematically they refer to the variability in the number of events per unit area for the whole point pattern. A second order moment quantity like the $K-$ function is a useful analytical tool for investigating second order effects in case of Gaussianity. Due to the non-Gaussian nature of aftershock epicenter patterns, however, second order moment measures are insufficient to fully describe these characteristics (Mateu, 2008) as clearly different processes may have identical second-order properties (Møller and Waagepetersen, 2007). As an alternative, nearest neighbor distribution functions such as the $F-$ and $G-$functions will be used. These are also formulated to measure the inter-point interactions but are defined as full probabilities.

Basic concepts and analytical methods are in e.g. Diggle (2003). Our interest concerns detection of systematics in the distribution, i.e. regularity or aggregation (clustersing) as deviation from randomness. Various summary statistics describing pattern properties can represent first order effects describing the number of epicenters per unit area, or

second order effects describing the dependency relationships between the epicenters.

Complete spatial randomness (CSR) is defined by the following criteria: (i) the number of epicenters in a region A of size $|A|$ follows a homogeneous Poisson distribution with mean $\lambda|A|$, where λ is the constant density; (ii) given n epicenters $x_i, i = 1, \ldots, n$ in A, the x_i are an independent random sample from the uniform distribution on A (Diggle, 2003). In other words, the density of the point pattern does not vary over the bounded region, and there are no interactions among the events.

Density estimation can be based on kernel functions Bowman and Azzalini (1997). Epicenters contribute to density estimation according to their distance from the kernel centre - the closer to the kernel centre, the larger the influence. The range of influence is limited by the kernel bandwidth controlling the smoothness of the result. Density plots with well-chosen bandwidth provide a good summary of the data, whereas a bandwidth that is too large leads to too much smoothing, and a bandwith that is too small over-emphasizes local epicenters that are not connected to the process that generates the aftershocks.

Dependency relationships for interactions can be described by nearest neighbor distances defined as the distance from the ith aftershock to the nearest other aftershock in A. The empirical cumulative probability distribution function $\hat{G}$ for the nearest neighbor distances summarizes the incident pattern as:

$$\hat{G}(w) = \frac{\sum_{w_i \leq w} 1}{n},$$

where w_i is a nearest neighbour distance for the ith aftershock and n is the number of aftershocks in A. The observed pattern, however, is usually part of a larger region, where the distribution of events is unknown. Interaction between events lying inside and outside the study region cannot be properly accounted for and cause edge effects. A simple but effective adjustment consists in reducing the sample by the buffer defined around the boundary. Aftershocks falling inside the buffer are not used for the analysis directly, but unveil the distribution behind the reduced study region.

We will distinguish large aftershocks and small aftershocks, depending on their magnitude. For an epicenter pattern consisting of these two types $x_i, i = 1, \ldots, n$ and $x_j, j = 1, \ldots, m$, $\hat{G_{i,j}}(w)$, the empirical cumulative probability distribution function of the distance from an earthquake of type i to the nearest earthquake of type j, equals

$$\hat{G}_{i,j}(w) = \frac{\sum_{w_{i,j} \leq w} 1}{n},$$

where $w_{i,j}$ denotes the smallest distances of an epicenter of an earthquake of one type to that of the other type. Second order effects, in particular modeled by the $\hat{G}_{i,j}(w)$ function provides insight into the relationships between type i (large) and type j (small) earthquakes. If there are considerably more nearest neighbors at short distances than what would be expected for random distribution, we can assume a correlation between epicenters. In this way, processes underlying aftershocks are unveiled that indicate the importance of particular exploratory variables.

To ease the interpretation, it is suitable to plot the $\hat{G}(w)-$function against the theoretical curve for CSR, which is (ignoring the edge effects):

$$G_{CSR}(w) = 1 - exp(-\lambda \pi w^2).$$

A visual inspection of the $\hat{G}(w)$ and $\hat{G}_{i,j}(w)$ may reveal the spatial distribution of the epicenters. Excess of nearest neighbors at short distances indicates clustering, whereas excess of long distance neighbors refers to regularity. Importance of the difference between $\hat{G}(w)$ and $G_{CSR}(w)$ is assessed by using Monte Carlo simulations. For this purpose, empirical cumulative probability distribution functions are generated for nearest neighbor distances for each of 99 realizations of a simulated CSR process with the same density as the original pattern. Its average provides a reference line, maximum and minimum values provide simulation envelopes.

2.2.3 The Strauss Point Process Model

The single-type Strauss process Let the earthquake-generating process be denoted X, the Strauss point process Strauss (1975) is defined by the density

$$f(X) = \alpha \cdot \beta^{n(X)} \cdot \gamma^{\,s(X)} \tag{2.1}$$

where β is the intensity of X, $n(X)$ is the number of occurrences and $s(X)$ is the number of pairs of epicenters which are not more than r units apart. The fixed parameter $r > 0$ denotes the distance around each epicenter within which this epicenter interacts with other epicenters. The parameter γ controls the strength of interaction between epicenters. If $\gamma = 0$ then the process is a hard core process i.e. the epicenters follow complete inhibition up to r and thus there is zero probability for any

two epicenters of X to lie closer than r units apart. For $0 < \gamma < 1$, the process exhibits inhibition (repulsion) between epicenters. The strength of inhibition is determined by closeness of γ value to zero. $\gamma = 1$ means that the epicenters do not exhibit any interaction and hence are independent. In this case the process reduces to a Poisson process with intensity β. The Strauss model was originally proposed as a model for clustering with $\gamma > 1$, but later it was discovered that the density [Eq.(2.1)] is not integrable as it results in infinity. Hence, the model gives an infinite number of points when the interaction parameter exceeds one and thus the Strauss model turned out to be a model for inhibition and is defined only for $0 < \gamma < 1$. Finally, α is the normalizing constant.

The conditional intensity $\lambda(u, X)$ of the epicenter generating Strauss process X determines the probability of finding an epicenter at a location u given the rest of the process, and it is given by:

$$\lambda(u, X) = \beta \cdot \gamma^{t(u,X)} \tag{2.2}$$

where $t(u, X)$ is the number of epicenters of X within a distance r of epicenter u. Figure 2.1 shows simulated realizations of a simple Strauss model in a square region of 10 units for different values of intensity β, interaction parameter γ and interaction radius r, given as Strauss (β, γ, r). The central figure is the Strauss$(1, 0.5, 0.5)$ process, the first column shows $\beta = 0.1$ and $\beta = 5$, the second column shows $\gamma = 0$ and $\gamma = 1$, and the third column shows $r = 0.01$ and $r = 1.6$. Clearly, the point density increases with an increase in β. For a small value of γ the process exhibits a hard core behavior with points surrounded by empty space, whereas shorter distances are occurring more often for the larger value of γ where the pattern tends to follow a Poisson process. Finally, with an increase in interaction radius (r) the point density of the Strauss process decreases.

The multi-type Strauss Process The multi-type Strauss model is an extension to the Strauss model. It can take clustering of a epicenters consisting of different types of earthquakes into account. The condition is that one of the relevant types is a hard core process Baddeley (2008). The hard core distance between the epicenters determines the level of clustering in the data Isham (1984). For a multi-type Strauss process consisting of two types of earthquakes $x_i, i = 1, \ldots, n$ and $x_j, j = 1, \ldots, m$, the conditional intensity $\lambda((u, i), X)$ describes the conditional probability of the occurrence of an earthquake of type i at location u,

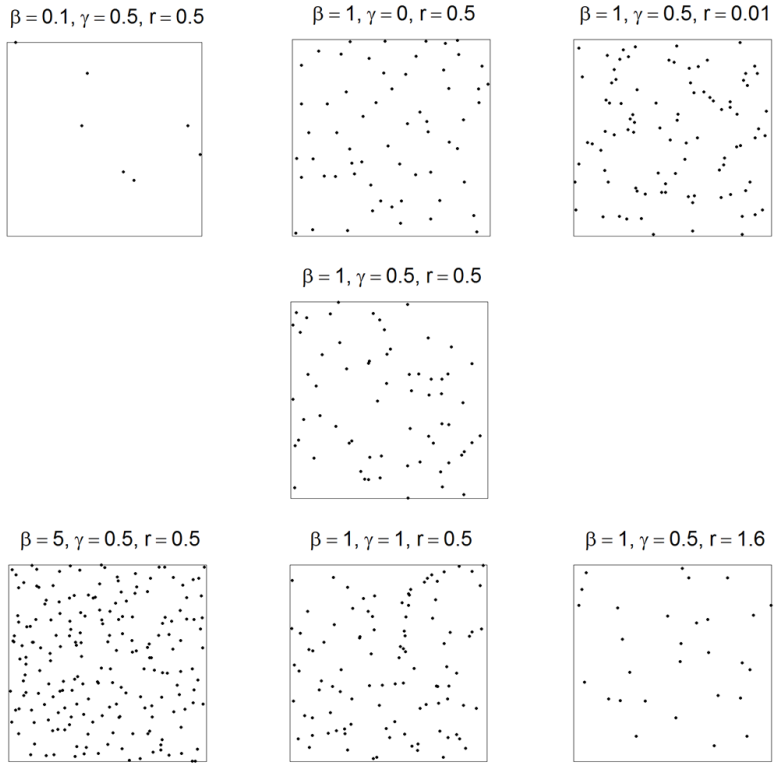


Figure 2.1: Realizations of a simple Strauss model within a square region of 10 units. The central figure has parameters (β,γ,r) = (1,0.5,0.5). The first column shows $\beta = 0.1$ and $\beta = 5$, the second column shows $\gamma = 0$ and $\gamma = 1$, the third column shows $r = 0.01$ and $r = 1.6$

and is given by

$$\lambda_i(u, X) = \beta_i \prod_j \gamma_{i,j}^{t_{i,j}(u,X)} \tag{2.3}$$

where $t_{i,j}(u, X)$ is the number of earthquakes in the marked point pattern of epicenters with mark equal to j, lying within a distance $r_{i,j}$ of epicenter u of an earthquake with mark i, and $\gamma_{i,j}$ are the interaction parameters for the pairs of epicenters of earthquakes of different types Baddeley (2008).

2.2.4 Model Diagnostics

We distinguish various formal and informal methods to check the goodness-of-fit of a fitted point process model Baddeley (2008). Formal methods include χ^2 goodness-of-fit tests and Monte Carlo tests, whereas the informal methods include the residual analysis of a fitted point process model Baddeley (2008); Diggle (2003). For Gibbs models χ^2 goodness-of-fit test and Monte Carlo tests are theoretically not supported. Instead, the common $K-$ and $G-$ functions are used to simulate the critical envelopes for the fitted models which are then used as tools for checking the validation of the fitted Gibbs models. The residual analysis Baddeley et al. (2005, 2008) serves as informal validation tool for Gibbs point process model where the residuals from linear regression are plotted against some explanatory variable Gelfand et al. (2010). If the fitted model is correct the residuals must have zero mean.

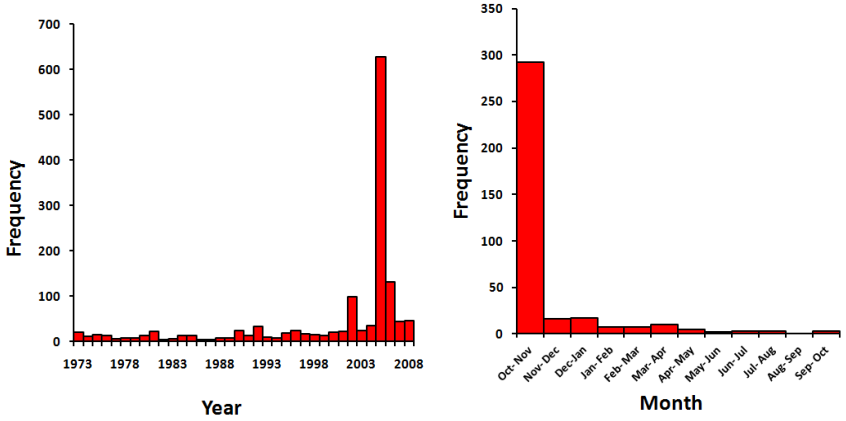
2.3 Earthquake Data

The data chosen for analysis and model fitting are the earthquake records of the Northern part of Pakistan. This region is an active seismic zone. It lies on the western edge of the Himalayan Arc, which denotes the region of continental convergence between the Indian and Eurasian tectonic plates. Location of tectonic plate boundary plays a significant role in determining seismicity of the study area. Compression motion between the two tectonic plates results in a series of large thrust faults including the Main Karakorum Thrust (MKT), the Main Boundary Thrust (MBT), and the Main Mantle Thrust (MMT) (Naranjo, 2008). Most of the aftershock of the Kashmir earthquake are concentrated around the north-western tip of the causative fault (Ömer Aydan, 2006). The data were taken from the USGS website (USGS, 2008). They include 1403 earthquakes that occurred in the region between January 1973 and August 2008.

Figure 2.2 presents the temporal analysis of the earthquake data. Figure 2.2a shows that a larger seismic activity in the region occurs in 2005 than in other years. This is due to the Kashmir $M_{7.6}$ earthquake, which struck the region on October 8, 2005, and was followed by a series of aftershocks.

For obtaining an insight into the spatial distribution of the aftershocks, we divided the total data set into earthquakes with magnitudes between 4 and 5.4 (both inclusive) that were marked as 'Small' and earthquakes with magnitudes above 5.4 as 'Large'. This division already shows a clear change in the seismic history of the area. For example, there were 22 $M_{\geq 5.5}$ earthquakes, of which 12 occurred at the same

day of the major earthquake and 3 more $M_{\geq 5.5}$ earthquakes occurred within 15 days after the major chock. Only 7 other $M_{\geq 5.5}$ earthquakes occurred during the past 35 years. We therefore focused on 2005 to analyse the spatial distribution of the aftershocks following the Kashmir earthquake.



(a) (b)

Figure 2.2: (a) Annual distribution of earthquakes in Northern part of Pakistan since 1973, and (b) Month-wise seismic record of the area for $M_{\geq 4}$ earthquakes since 1973

Figure 2.2b shows the month-wise seismic record of the area for $M_{\geq 4}$ earthquakes since October 8, 2005. The figure shows that seismicity of the area decreases after the first month and that the number of events in the preceding months is almost negligible as compared to the number of earthquakes in the first month after the main shock. We, therefore, focused on the earthquake data of the first month (Fig.2.3).

Figure 2.3 shows that only four epicenters lie beyond 50 km of the aftershocks region and hence do not seem to be a part of the seismicity resulting from the Kashmir earthquake. These earthquakes of magnitudes 4, 4.1, 4.5 4.8, respectively, were all less than 5.5, taken as a lower threshold value to define a large earthquake. We ignored these earthquakes and limited the study area to the part where the aftershocks occurred within one month after the October 8, 2005. Since there is no naturally defined boundary of the aftershocks region, a rectangular boundary with coordinates $(274963, 3849631)$, $(348760, 3722997)$, $(445264, 3779328)$ and $(371467, 3906398)$ in WGS-84 with projection system UTM zone 42 North was used as the boundary of the study region (Fig. 2.3).

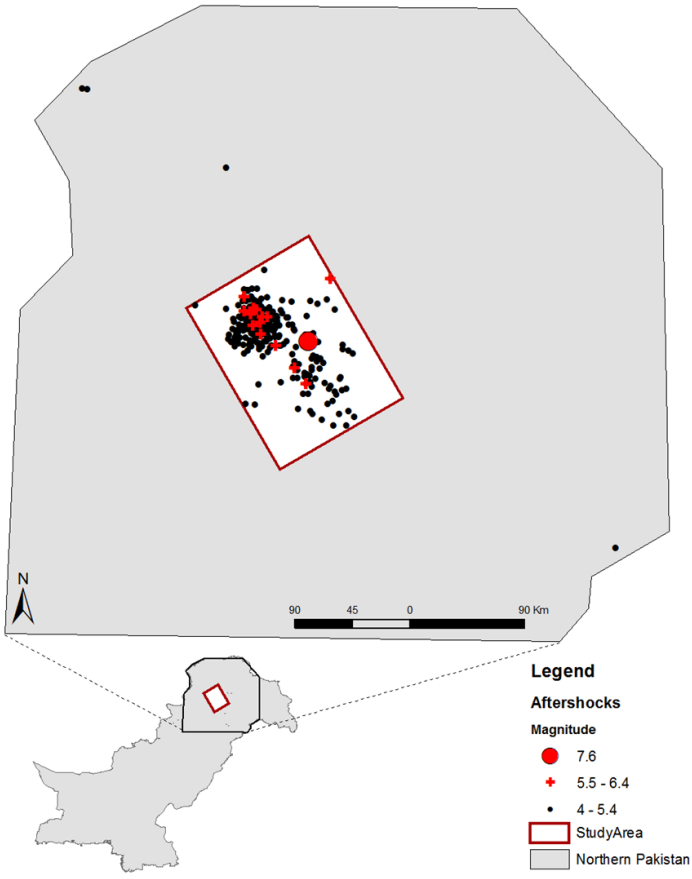
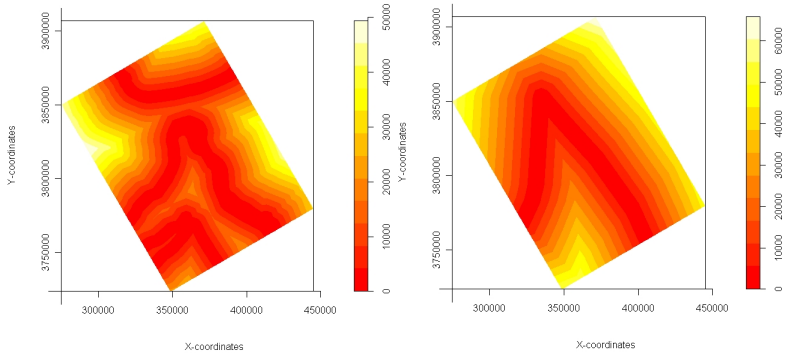


Figure 2.3: Aftershocks of Kashmir earthquakes within one month

Geological faults and plate boundaries constitute the additional information. They carry information on geological phenomena that may affect the spatial occurrence of earthquakes. To assess the influence of geological faults and the plate boundaries located in the study region on the earthquakes distribution pattern, two pixel images (Fig. 2.4a and Fig. 2.4b) were created, showing shortest distances of all epicenters from the nearest fault location (DF) and from the plate boundary (DP) respectively.



(a) (b)

Figure 2.4: Pixel images of (a) distance from the geological faults, and (b) distance from the plate boundary

2.4 Results

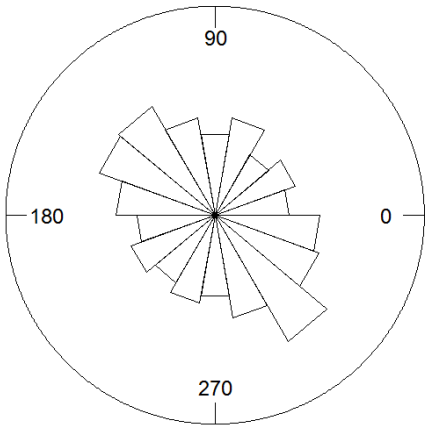
2.4.1 Directional analysis

For assessing the directional characteristics of the earthquakes data relative to some important geological locations, rose diagrams were drawn for the data converted to directional type (Fig.2.5).

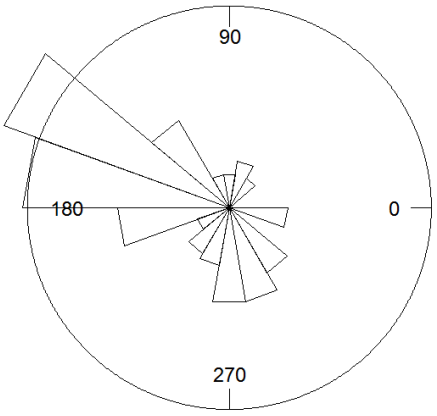
The rose diagram (Fig. 2.5a) shows the angles between each earthquake epicenter location and the location of the most recent prior earthquake. Such a graph should allow us to detect a possible spatial path taken by the earthquakes with respect to time of occurrence. The rose diagram shows an almost uniform distribution, with some preference to the NW and SE directions. Apparently the aftershocks follow each other in this direction with similar frequencies into both directions.

Orientations of earthquake epicenter locations with respect to the epicenter of the main shock of Kashmir earthquake are shown in Figure 2.5b. The diagram shows a high concentration of earthquakes between 130^o and 190^o of the epicenter of the main shock and another cluster of points between 260^o and 310^o. From Figure 2.5b we observe that most earthquakes are closely concentrated around the convergence point of the two plate boundaries. It shows that the convergence point to a large degree determines the occurrence of earthquakes.

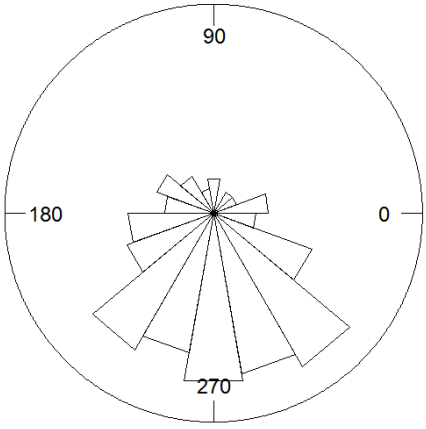
Next, we determined the angles of all aftershock epicenters with respect to the convergence point of the two plate boundaries (Fig. 2.5c). This is to test the main characteristics of the earthquake data considered relative to this point. The rose diagram suggests a strong clustering of



(a) Angle of each earthquake epicentre w.r.t. its previous earthquake location shows occurrence of earthquakes almost uniformly in all directions with time



(b) Angles of all earthquake epicentres w.r.t. the main shock showing a major cluster of aftershocks in the north-west direction



(c) Angles of each earthquake epicentre w.r.t. the convergence point of the plate boundaries showing most of the aftershocks clustered in the South

Figure 2.5: Rose diagrams for the data converted to directional type

earthquakes epicenters to the south of this point.

2.4.2 First order and second order effects

Epicenter intensity was estimated and plotted using kernel smoothing (Baddeley, 2008; Turner, 2009a) (Fig. 2.6). Issues involved in suitable selection of bandwidth for kernel smoothing are discussed in detail in Illian et al. (2008). After trying a number of values, bandwidth was finally selected to be 10% of the diameter of a square with area equal to size of the study area as it gave an appropriate generalization of the variability in intensity of earthquake occurrences. Epicenter intensity is heterogeneous throughout the study area as Figure 2.6 shows a high epicenter concentration in the north-western part of the study area, making it a hot-spot for aftershock occurrences.

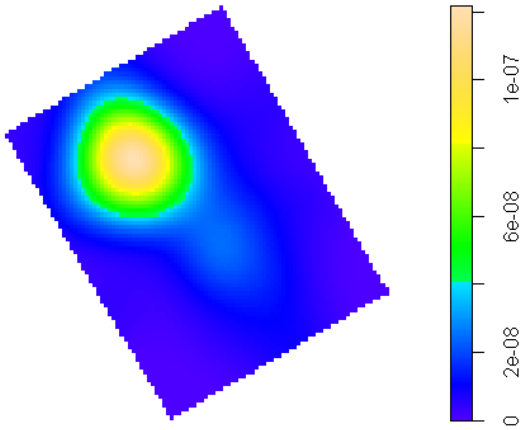
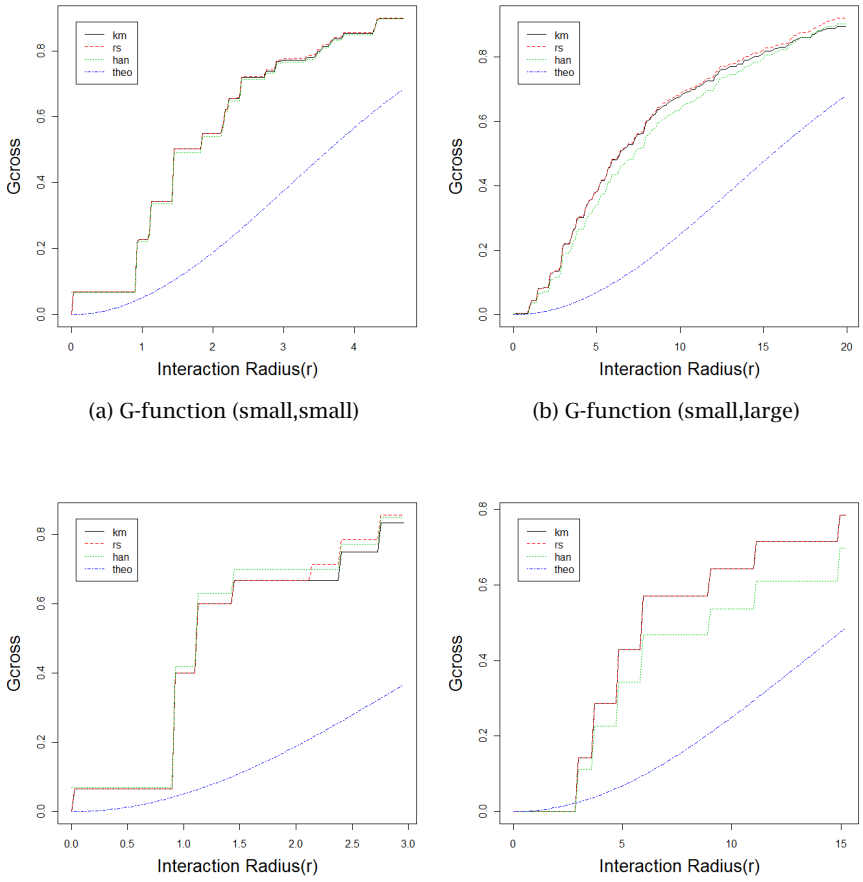


Figure 2.6: Epicenter intensity of the Kashmir earthquake aftershocks

Second order effects were investigated using distance based $G-$ functions for multi-type marked patterns.

Figure 2.7 shows the $G-$ function between different types of earthquakes. Since the estimated curves are above the theoretical curves of the Poisson process, the epicenter pattern exhibits clustering. In Figure 2.8a and 2.8c, we see nugget effects of the curves, which is due to 21 coincident epicenters at 10 different locations. The presence of coincident epicenters violates the basic assumption of the point pattern theory that all the points should be distinct. As from the $G-$functions (Fig. 2.7) we observe no earthquakes within a radius of 1 km from the coincident locations; to circumvent the violation we kept the epicenters with highest magnitudes at their original positions but introduced uniform random disturbances $U(0, 0.5)$ to both coordinates of other coincident epicenters in order to disturb them from their original loca-



(a) G-function (small,small)

(b) G-function (small,large)

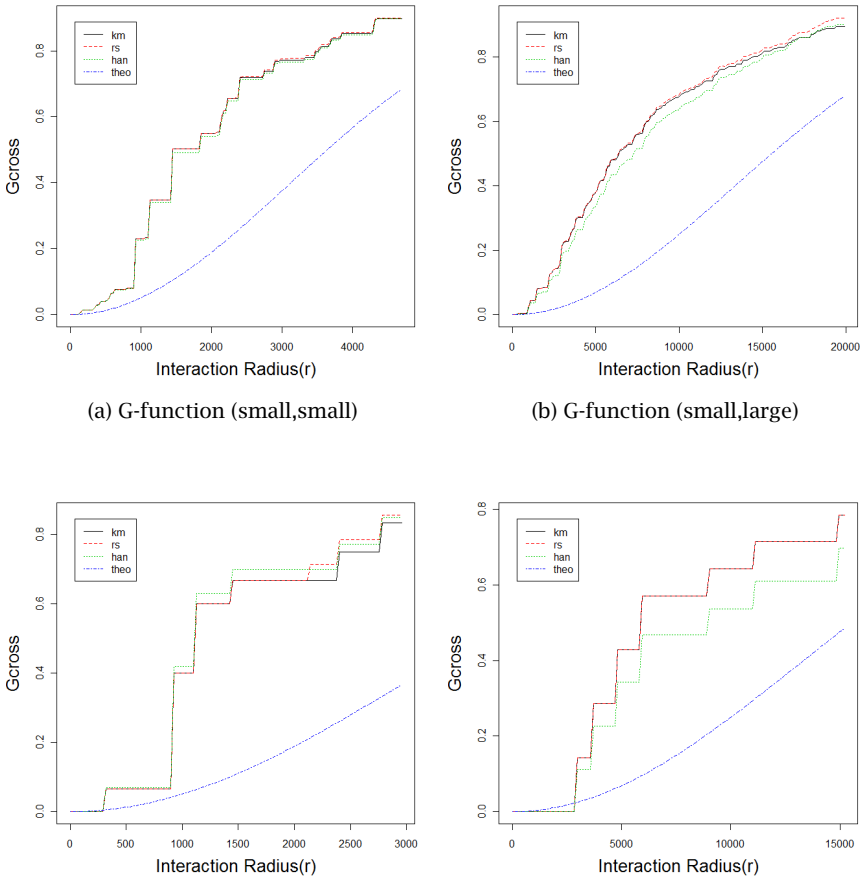
(c) G-function (large,small)

(d) G-function (large,large)

Figure 2.7: Nearest neighbour $G-$ functions for pairs of different types of earthquakes (the interaction radii here are given in kilometers(km))

tions. The $G-$functions were again calculated for the modified data and the results are shown in Figure 2.8.

Figure 2.8 shows that the above described strategy did not influence the results and the overall pattern except that the nugget effect is now removed.



(a) G-function (small,small)

(b) G-function (small,large)

(c) G-function (large,small)

(d) G-function (large,large)

Figure 2.8: Nearest neighbour $G-$ functions for pairs of different types of earthquakes (modified)(the interaction radii here are given in meters(m))

2.4.3 The Strauss Model

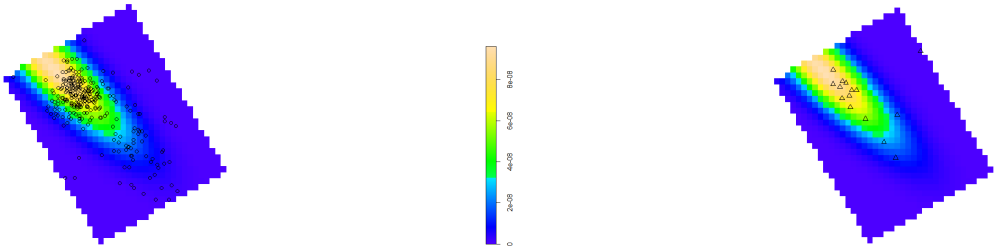
As the above $G-$functions reveal clustering of the earthquakes and the $G_{L,L}$ function shows a hard core distance of approximately 3 km, we selected the multi-type Strauss model for data fitting. Estimation of interaction radii between earthquakes of different types, however, remains unclear in point pattern theory. From visual inspection of the above plotted $G-$functions, we obtained an initial rough estimate of the

interaction radii and then several iterations of the multi-Strauss model for different values of r were made, and the following matrix for the interaction radii was finally determined.

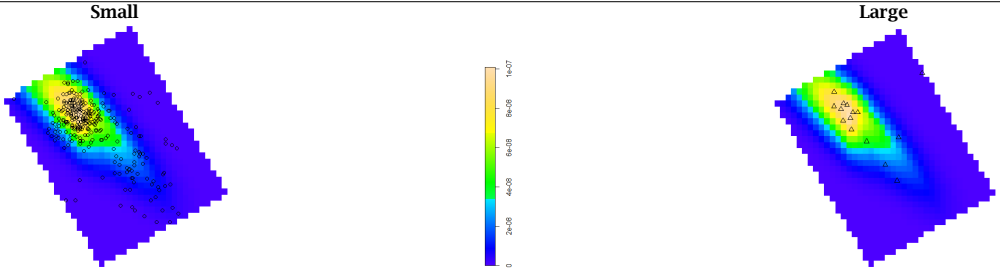
$$r_{i,j} = \begin{bmatrix} 1000 & 10000 \\ 10000 & 20000 \end{bmatrix}$$

Note that the matrix for interaction radii must be symmetric as Gibbs models assume symmetric interactions between points. Although G functions were visually analyzed to get some idea of the interaction distances, radii for the earthquakes pairs had to be adjusted in order to bring the matrix to symmetric form. Different forms of intensity function for the Strauss model were fitted by incorporating the available geological information one by one. Table 4.1 shows the log intensity functions for all the models, the trends fitted by them to 'Small' and 'Large' earthquakes, along with their respective AIC values.

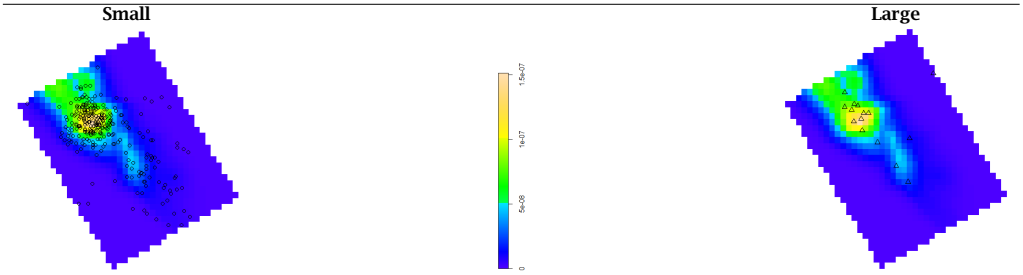
Model 1: $\log b(x, y) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 xy + \beta_5 y^2$
AIC = 9116



Model 2: $\log b(x, y, DP) = \beta_0 + \beta_1 DP + \beta_2 x + \beta_3 y + \beta_4 x^2 + \beta_5 xy + \beta_6 y^2$
AIC = 9113



Model 3: $\log b(x, y, DF) = \beta_0 + \beta_1 DF + \beta_2 x + \beta_3 y + \beta_4 x^2 + \beta_5 xy + \beta_6 y^2$
AIC = 9093



Model 4: $\log b(x, y, DP, DF) = \beta_0 + \beta_1 DP + \beta_2 DF + \beta_3 x + \beta_4 y + \beta_5 x^2 + \beta_6 xy + \beta_7 y^2$
AIC = 9090

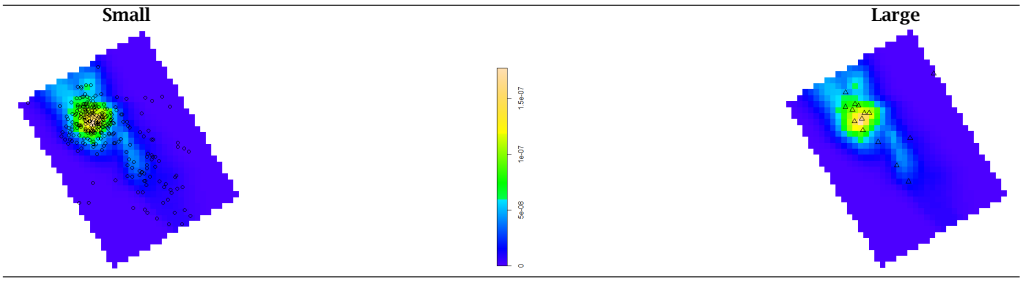


Table 2.1: Trends fitted by the models to 'Small' and 'Large' earthquakes, along with their AIC values. AIC values decrease with inclusion of covariates.

Table 2.2 shows estimated parameters of fitted intensities for the different models. Covariates DF and DP represent the shortest distances of all epicenters from the nearest fault location and from the nearest plate boundary respectively.

	b	x	y	x^2	xy	y^2	DP	DF
	$\times 10^4$	$\times 10^{-2}$			$\times 10^{-9}$		$\times 10^{-5}$	
Model 1	-3.08	1.28	1.49	-1.69	-3.04	-1.81		
Model 2	-2.72	1.14	1.32	-1.43	-2.74	-1.60	-3.22	
Model 3	-3.40	1.69	1.62	-2.50	-3.96	-1.93		7.97
Model 4	-2.95	1.52	1.40	-2.18	-3.57	-1.66	-3.06	7.84

Table 2.2: Estimated parameters of the fitted intensity

Table 2.3 shows the estimated values of interaction parameters. We observe that with inclusion of more covariates, values of different interaction parameters converge to one indicating an absence of interaction i.e. the earthquake process tends to be a Poisson point process. If all of the variability could be explained in terms of covariate effects, the interaction among points becomes negligible and the point pattern exhibits characteristics of a Poisson point process.

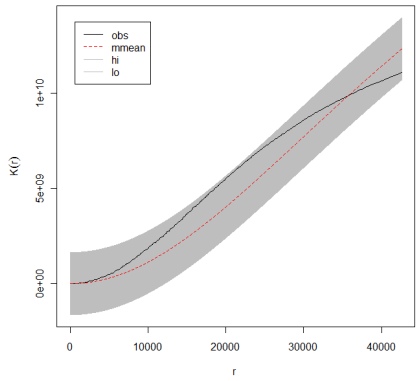
	Small-Small	**Small-Large**	**Large-Large**
Model 1	1.370	1.156	0.282
Model 2	1.349	1.139	0.310
Model 3	1.237	1.126	0.326
Model 4	1.230	1.113	0.352

Table 2.3: Estimated interaction parameters

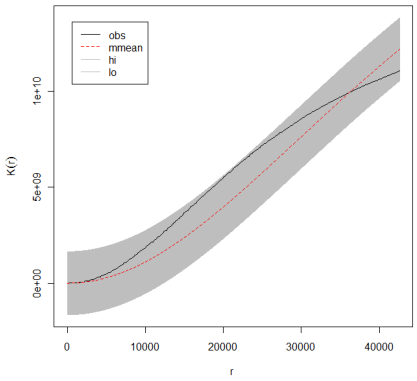
For checking goodness- of -fit of all the fitted models, critical envelopes were simulated using the K summary function (Fig. 2.9) which show that the fitted models fit sufficiently well to the earthquake pattern.

2.5 Discussion

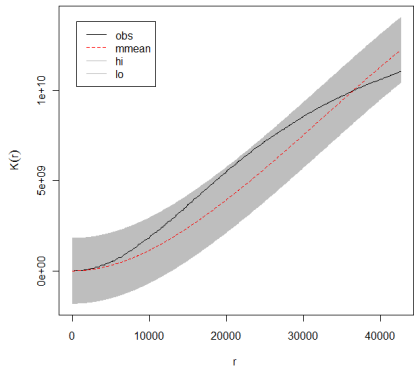
For the present study potential of the Strauss point process was investigated for capturing variability in the point pattern consisting of locations and magnitudes of the earthquakes following the Kashmir earthquake 2005. This model allows the use of covariates for explaining spatial variations in a marked point pattern where the marks can be categorized as factors. The Strauss model proved to be flexible and rigorous in modelling the clustered pattern of the epicenters and it



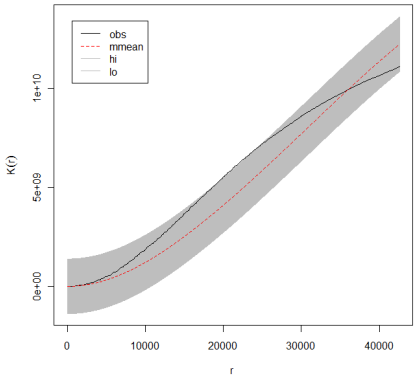
(a) Simulated envelopes for Model 1



(b) Simulated envelopes for Model 2



(c) Simulated envelopes for Model 3



(d) Simulated envelopes for Model 4

Figure 2.9: Nearest neighbour $G-$ functions for pairs of different types of earthquakes

could explain the variability of epicenter distribution. With the inclusion of explanatory variables, the results of Strauss model fitting became increasingly refined; hence it has the capacity to incorporate a variety of additional variables to define a marked point pattern such as that of epicenters of aftershocks of different types.

The earthquake data serve as an example of marked point pattern, difficult to fully explain since the variability or in-homogeneity in the

pattern is caused by many observed and unobserved geological and geophysical factors. Earthquakes relate to subsurface physical processes and their generation is spanned over long periods of time and depends on several forces interacting in the earth's crust. For a realistic model fitting, more and more geological and geophysical information is required. For the earthquake data, there are only a few observable factors determining the epicenters as earthquakes are generated deep below the earth's surface. Phenomena happening at such depths may or may not manifest itself in the form of clear indications above the earth's surface and this puts one into difficulty in searching for possible indications or covariates which could be associated with the spread of earthquake locations. Some of the geological factors considered as being highly associated with the epicenters were incorporated in the Strauss model. The modelling approach suggests that the epicenters close to the plate boundaries are associated with a higher earthquake probability as can be observed from Figure 2.5c. The large clustering of epicenters at this location might be due to the fact that this location is under effect of two close-by plate boundaries. Locations of active tectonic faults also proved to be a significant determinant for the distribution of earthquake epicenters as incorporation of the information about the fault locations improves the model fitting results. Finally when the effects of both the plate boundary locations and geological fault locations are considered as explanatory variables, model fitting further improves in terms of the fitted intensity (Table 4.1) and also in terms of the AIC. If information about more geological or geo-physical factors could be available and incorporated in the Strauss model fitting, the model is expected to be even better representative of the earthquake locations distribution in the study region.

Despite the limitations and complexity of the data, some limitations of the Strauss model were encountered during the study. Basically the model serves as a model of inhibition between the points of a point pattern and it accounts for clustering only when the data are converted into a multi- type point pattern. For earthquakes data, it can be difficult to split the marks (magnitudes for earthquakes) into a number of categories; as defining and reasoning thresholds for different factors could be vague. The other condition for taking clustering effect of a point pattern into account by Strauss model is that one of the categories of the multi-type pattern keeps a hard core distance, which again in some cases may not be observed in reality and in that case the Strauss model cannot be applied as a model of clustering.

Another assumption of the point process theory is that all data points are non-overlapping. This puts one into question as to what to do if the data set consists of some overlapping points as in that case several of the formulae and interpretations of the point process theory

become invalid and one has no choice except to manipulate the data in order to get rid of the duplication. Identical points encountered in the current study are distributed randomly over a small distance to overcome the problem. Spatial statistics, however, lacks the modelling techniques for a data set consisting of many overlapping points.

The Strauss model, and more generally the Gibbs models, assumes symmetric interactions between the points of a pattern which might not be the case in reality as for earthquakes data the locations of major earthquakes affect the locations of small earthquakes but not vice versa, or in other words, the locations of large earthquakes are determinant of small earthquakes in the form of aftershocks but small earthquakes do not have the same effect on the occurrence of large ones. For a realistic model fitting, geological reasoning cannot be ignored and the model should be capable of taking into account the non-symmetric interactions.

The estimation of interaction radii is another tricky issue in spatial point process modelling for a multi-type marked data as there are no hard and fast rules to determine the interaction radii from the data set. One needs several iterations of model fitting by using different values of interaction radii in order to come up with the values that fit a model best representing the point locations. For our analysis, an initial rough estimate of the interaction radii for multi-type Strauss model was drawn from the $G-$functions and several iterations of the Strauss model were done and the results compared to select the final matrix of interaction radii. This practice is time consuming and may not be fully reliable; however, there is no technique to estimate the values directly from the data points.

Generally the point pattern approach requires the spatial covariates to be spatially continuous i.e. the values of the explanatory variables must be available for each and every point in the study region (pixel image) or at least at some other locations apart from the data points (dummy variables). The point process theory needs to be modified such that the explanatory variables which are available only for the data locations (e.g depth) can also be utilized in model fitting.

2.6 Conclusion

Application of the Strauss point process model to earthquake aftershock data proved satisfactory in explaining the spatial trends and capturing the sources of variability. The explanatory variables consisted of the information about the spatial location of the plate boundaries and geological faults in the study area. The study addressed those as pixel images showing shortest distance to the nearest plate boundary and

nearest fault location for each pixel. The effects of the explanatory variables were quantified by improvement in AIC values. Locations of plate boundaries and geological faults are significant determinants for the position of earthquake epicenters. Combining the effects of both these variables with the magnitudes and geographic locations of the earthquake epicenters, we were able to significantly improve the modelling.

Detection of selective logging as a forest LULC change

3

Abstract

The Brazilian Amazonian rain forests are under imminent threat of serious degradation and ultimately deforestation. Human activities such as selective logging are an important cause. Selective logging locations are difficult to detect from medium-resolution Landsat images, due to their relatively small sizes and subtle spatial patterns. Spectral linear unmixing provides an effective tool for the purpose. The orientation of geometrically corrected images, however, artificially introduces zero-reflectance background pixels. These change the variance–covariance structure of the image bands and hinder the identification of pure endmembers. In this study we compare image cropping and image rotation as two alternative approaches. Selectively-logged forests were detected in the northern Rondonia state, north-western Mato Grosso state, and south-eastern Amazonas state in Brazil by applying spectral unmixing. The study shows that image rotation is a better approach as it preserves the image extent and thus provides information on forest degradation over a wider region. Spatial statistical analysis of the detected locations shows a strong clustering within the study area. We conclude that the endmembers used in this study represent basic components of a degraded forest environment. As spectral unmixing of remote sensing images avoids collection of field data, it may broadly be applied towards other Amazonian regions as well.

[1]This chapter is based on the following paper

Anwar, S. and Stein, A., 2012. Detection and spatial analysis of selective logging with geometrically corrected Landsat images. *International Journal of Remote Sensing* 33, 7820-7843.

3.1 Introduction

The Brazilian Amazon is the world's largest contiguous rain forest. It covers about one third of the world's remaining tropical forest with an area approximating 4.1 million km^2. This forest is under a constant threat of depletion due to accelerating encroachments by humans. Since the 1970's, approximately 16% of its original rainforest has been destroyed due to anthropogenic influence (Aguiar et al., 2007). The average annual deforestation rate for the period 2001–2010 is estimated to be 16531 km^2 – more than half of the total extent of Belgium (30510 km^2) (INPE, 2011a). The current deforestation rate bears the potential of driving strong environmental consequences if it continues unchecked (Laurance et al., 1998).

In the Brazilian Amazonia, selective logging for timber is a major source of forest degradation (Uhl et al., 1991). It causes long-lasting deteriorating effects to forest phenology by damaging the canopy structure and its floristic composition. This may disturb functioning of the entire forest ecosystem by jeopardizing the survival of other associated species (Koltunov et al., 2009). Selective logging also increases the vulnerability of a forest to fires by providing fuel in the form of debris left after logging (Gerwing, 2002). The canopy gaps resulting from selective logging allow sun and wind to create understory dryness favourable for spread of fire (Fearnside, 2005). Moreover, selective logging causes forest fragmentation (Broadbent et al., 2008) which affects the diversity of forest fauna. The skid trails and log decks built for logging activities provide access to interiors of the vast Amazonian forests to exploit it for cattle ranching or agricultural productions, thus leading to wide spread of deforestation (Uhl et al., 1991; Fearnside, 2005).

Accurate and precise detection of selectively-logged sites in a forest is crucial for analysing the spatial distribution of forest degradation and deforestation in the Amazonia. Remote sensing, due to its synoptic view and fast coverage, may serve as a viable source for repeated mapping of dense and inaccessible vast areas of the Amazonian forests. It can be used to monitor selective logging over Amazonia, which otherwise requires labor-intensive and time-consuming field surveys. Selective logging is, however, not easily identifiable through satellite images due to its spatial pattern which is often less than the remotely-sensed image resolution. Interpretation of remote sensing data for selective logging detection is, therefore, a challenging task and requires special image mining techniques (Gerwing, 2002).

Several different techniques using satellite data have been implemented for detection and quantification of areas under selective logging, ranging from a simple visual interpretation of the satellite imagery

(Stone and P., 1998; Matricardi et al., 2007) and texture analysis (Asner et al., 2002), to the more technical spectral mixture modelling (Souza and Barreto, 2000; Santos et al., 2003; Souza et al., 2003).

Spectral mixture modelling (Shimabukuro and Smith, 1991) of multispectral or hyperspectral imagery serves as a suitable approach for detection of selective logging (Adam et al., 1995; Lu et al., 2003; Souza et al., 2003; Lu et al., 2004; Souza Jr. et al., 2005; Matricardi et al., 2010). Soil fraction images obtained by means of spectral mixture models allow identification of small forest perforations with spatial patterns depicting at sub-pixel scale in remote sensing imagery (Souza et al., 2003).

Geometric correction of nonsystematic distortion is a vital preprocessing step for remote sensing images in any practical application of remote sensing data. Nonsystematic distortions are introduced due to the earth rotation during image acquisition and the satellite orbit inclination. To correct for earth rotation effect, the image scan lines are offset towards the west, resulting in a parallelogram outline of the image. The Landsat-5 satellite has an orbit inclination angle of 8.2^o, thus each acquired scene represents a parallelogram with this orientation angle. The orientation angle measures the anti-clockwise rotation to produce a north-oriented image. Rotation may help to more accurately represent features on the ground. To maintain the raster (matrix) format of the image after rotation, an artificial background with values set to zero is added to the image, which is treated as a part of the image during further image processing (Jensen, 2004). The artificial background adds a large number of pixels with zero reflectance to the original image. These pixels affect the variance–covariance structure of the image bands. Estimation of the parameters in spectral mixture modelling, in the presence of a large number of background pixels, gives erronous results. This makes linear spectral unmixing technique ineffective to be applied with oriented images.

Logging operations vary in timing, location and intensity (Matricardi et al., 2005). A spatial statistical analysis of the detected logged-locations may reveal important spatial characteristics of selective logging distribution throughout the Amazonia. Quantitative assessment of a spatial pattern consisting of small-scale disturbances, such as canopy perforations due to selective logging as reflected in satellite imagery, is important to understand large-scale forest dynamics (Shimatani and Kubota, 2004). Spatial analysis of logging distribution in the Amazonia may also be helpful in determining the most important factors responsible for the Amazonian forests degradation. Ultimately, this can assist in formulting effective strategies for conservation and exploitation of the forest resources (Santos et al., 2003).

This study aimed at detecting changes in forest canopy due to selective logging by eliminating the influence of zero-reflectance background

introduced by image orientation. Two strategies for elimination of the background are proposed: (1) cropping the image to include only the area covered by a rectangle inscribed by the true image, (2) removing the image orientation by using the orientation angle specified in the image meta-data file. For detection of selective logging, the methodology was applied to Landsat5 Thematic Mapper (TM) scene (row 230, path 66), covering parts of Mato Grosso, Rondônia and Amazonas states. A spatial statistical analysis of the detected selectively-logged sites was also carried out.

3.2 Data description

A Landsat5-TM image of the study area (described in Chapter 1), with row 231 and path 66, acquired on 21 July, 2008 was used in this study. The image was downloaded from the Brazilian Space Agency (INPE) website (www.dgi.inpe.br). The website offers processing level-2 products, which means that the images are radiometrically and geometrically corrected. This is the highest pre-processing level applied to the Landsat images at INPE, with positional error of the images upto 1 km (Vergara, 2009). The image did not exhibit any atmospheric contamination like haze or clouds, therefore it was converted to top-of-atmosphere reflectance without performing atmospheric correction.

For checking accuracy of the results, cloud free images of CBERS-2B satellite, sensor HRC (panchromatic, resolution 2.5 m) were downloaded from the INPE website. The website offers georeferenced CBERS images. The georeferencing is done without the use of control points and therefore the images may have positional errors ranging from hundreds of meters to few kilometers (INPE, 2011a). An extrenal source of georeferencing, therefore, was needed to eliminate the positional errors. Only those images which covered the logged forest areas were selected for the accuracy assessment (Fig. **??**). The georeferenced Landsat image was used as a reference image to register the CBERS images using the ENVI software. The maximum root-mean-square error was recorded as 15–20 CBERS pixels.

3.3 Methodology

3.3.1 Elimination of Image background

A large number of zero-reflectance background pixels in a geometrically-corrected Landsat image, apart from providing misleading information about the bands variability, also causes poblems in identifying the image

endmembers using the automated Pixel Purity Index (PPI) (Section 3.3.5) and the $p-D$ visualiser (Section 3.3.6). These zero-reflectance pixels are identified by the PPI as the purest pixels in the image. The $p-D$ visualiser, which loads only the purest pixels as its input, ignores the true image data and hence the image endmembers identification using the automated procedure described above does not remain possible. To eliminate influence of the background pixels on spectral unmixing, an image has to be cropped to include only the area covered by a rectangle inscribed by the true image before spectral unmixing of the image can be carried out. As an alternative, the image orientation can be removed by rotating the image anti-clockwise to the orientation angle specifed in the image meta-data file. The surrounding dark background area can then be removed by cropping the image into a rectangular region consisting of only the area covered by the true image. After spectral unmixing, the resultant images have to be rotated back to the original orientation. Selectively logged sites can then be detected from both cropped and rotated images, and results of the two strategies can be compared at the end.

3.3.2 Linear spectral unmixing

Linear spectral mixture for remote sensing images assumes that the spectrum of a pixel recorded by a sensor is created by a linear combination of the spectra of all component materials within the pixel (Adam et al., 1986; Smith et al., 1990; Shimabukuro and Smith, 1991; Rencz, 1999). By linearity we mean that no significant amounts of multiple scattering exist between the different component materials on ground and that each photon that reaches the sensor has interacted with only one component material.

In a degraded forest environment, four types of component materials are expected; green vegetation, non-photosynthetic vegetation, soil and shade. Linear unmixing techniques can be applied on Landsat images for finding the relative abundance of each of these materials in a pixel, using its bands 1,..,5, 7. Band 6, which is a thermal band and hence does not contain information meaningful for vegetation studies, can be excluded. The linear spectral mixture approach can then be expressed as

$$R_i(x) = \sum_{j=1}^{m} F_j R_{i,j}(x) + \epsilon_i(x), \tag{3.1}$$

where

$i = 1, ..., (n = 6)$ is the spectral band of the Landsat image,

$j = 1, ..., (m = 4)$ is the component material or endmember expected to be found in a degraded forest environment,

$R_i(x)$ is the spectral reflectance at pixel x in band i,

$F_j(x)$ is the proportion of endmember j within the pixel x, with $0 \leq \sum_{j=i}^{m} F_j \leq 1$,

$R_{i,j}(x)$ is spectral reflectance of endmember j at pixel x in band i , and

ϵ_i , the residual error, is the difference between the measured and modeled reflectance in band i, thus representing the part of a pixel spectrum not taken care of by the endmembers.

Linear spectral unmixing aims at solving the equation for the F_j, i.e. for finding the relative proportion of green vegetation, non-photosynthetic vegetation, soil and shade within each pixel of a Landsat scene. It estimates the reflectance for each individual endmember from the total reflectance of a pixel using matrix inversion.

To solve equation 3.1 for F_j for any application, the following prerequisites should be met: (1) $m \leq n$, (2) selected endmembers should not be a linear combination of each other, and (3) the spectral bands selected for analysis should not be highly correlated (Van Der Meer and De Jong, 2000).

For the four endmembers, linear spectral unmixing results in four fraction images, one for each endmember, and an RMSE (root mean square error) image to represent the residuals ϵ_i. Each fraction image provides pixel-wise the proportional abundance of an endmember, i.e. the areal proportion of the endmember in each pixel. Unmixing thus results in quantitative, and not in thematic classes (Settle and Drake, 1993), and therefore it provides a physical meaning to the remotely sensed spectral data (Souza et al., 2003). The RMSE image serves as an indicator of the suitability of the selected endmembers. Ideally, if the pixels only contain the selected endmembers, the RMSE image reflects complete noise (Van Der Meer and De Jong, 2000). Clearly visible RMSE patterns are indicators of missing or inaccurately selected endmembers (Lu et al., 2003).

In this analysis, spectral unmixing of Landsat image was carried out using the 'Spectral Hourglass Wizard' tool provided by the ENVI software.

3.3.3 Selection of endmembers

Traditionally, three to four endmembers are used in spectral unmixing performed for detection of selective logging. For example Souza and Barreto (2000); Monteiro et al. (2003); Lu et al. (2003) and Lu et al. (2004) used three endmembers representing green vegetation, soil and shade, whereas Adam et al. (1995); Souza et al. (2003) and Matricardi et al.

(2010) included one more endmember given by Non-photosenthetic Vegetation (NPV), which represented grass, leaf litter and woody material. Roberts et al. (1993) observed that more than 98% of the spectral variability in a forest could be explained by a linear mixture of three endmembers: green vegetation, shade and soil. For Landsat TM, NPV and soil exhibit minimum spectral contrast and in the absence of detailed field knowledge, it is difficult to distinguish between the two spectral classes (Adam et al., 1995). Moreover, Roberts et al. (1993) observed that when NPV and soil were both included as endmembers, the residual error was not significantly reduced and there was an increase in the uncertainty (noise) in the endmembers fractions. Therefore three endmembers; green vegetation (Veg), shade (Sh) and soil (So), are used in this study. Endmembers for these classes were determined by means of scatter plots of minimum noise fraction (MNF) transformed bands, pixel purity index (PPI) image (Section 3.3.5), $p - D$ visualiser (Section 3.3.6), and examination of the RMSE image.

3.3.4 Minimum Noise Fraction (MNF) transform

Presence of intercorrelation among explanatory variables in a regression model results in unstable estimates of the regression coefficients. Correlated explanatory variables exhibit little variation, making it difficult to disentangle the separate effects of each of the explanatory variables on the dependent variable. Solution of the regression equations using matrix inversion also requires the explanatory variables to be strictly orthogonal, as the matrix cannot be inverted in case of interdependence among explanatory variables. In multispectral data such as Landsat, however, often the bands are highly correlated due to low spectral resolution of the sensors. Solution of equation 3.1 turns out to be errenous in presence of high between-band correlation. Principal Components (PCs) help to compress the image information by removing the between-band correlation before solving them. The compression results in a steadily decreasing signal-to-noise ratio as the number of PCs increases, making it possible to separate noise from signal. As observed by Green et al. (1998), the standard PC transform does not always result in new image bands that show steadily decreasing image quality with an increasing number of PCs. To solve this problem, the Minimum Noise Fraction (MNF) transform has been developed (Green et al., 1998) and applied in this analysis, which performs the optimal ordering of the image bands in terms of quality (Green et al., 1998).

The MNF transform considers a multivariate data set of n bands with intensity levels $R_i(x), i = 1, ..., n$; where x gives the coordinates of the sample. It assumes that

$$R_i(x) = S_i(x) + N_i(x),$$

where $S_i(x)$ and $N_i(x)$ are the uncorrelated signal and noise components of $R_i(x)$. Thus, the covariance matrix of $R_i(x)$ equals

$$Cov\{R_i(x)\} = \Sigma = \Sigma_S + \Sigma_N,$$

where Σ_S and Σ_N are the covariance matrices of $S_i(x)$ and $N_i(x)$, respectively. The MNF transform aims at maximizing the signal-to-noise ratio given by

$$Var\{S_i(x)\}/Var\{N_i(x)\},$$

using the linear transformations

$$Z_i(x) = a_i^T(x)R_i(x), i = 1, ..., n.$$

this is done such that the signal-to-noise ratio for $Z_i(x)$ is maximum among all linear transformations orthogonal to $Z_j(x), j = i, ..., n$. Here a_i are the eigenvectors of Σ_N and Σ^{-1}, and the eigenvalue corresponding to a_i equals the signal-to-noise ratio in $Z_i(x)$.

The MNF transform sets the noise variance to 1 in all bands. Noise is estimated from the image by the shift difference method. This method is built on the fact that, in most remote sensing data, the signal at a pixel in the image is strongly correlated with the signals at neighboring pixels, whereas the noise shows only a weak spatial correlation. The noise image is obtained by subtracting line-shifed and sample-shifted images from the original image and averaging and scaling the two results. The MNF tranformed bands are ranked with the largest variance in the first band and decreasing variance with increasing band number until noise remains. The bands containing only noise are not used in the subsequent processing whereas the bands with larger variances are examined to determine if they constitute useful information (Bateson and Curtiss, 1996). For Landsat TM data, more than 90% of the spectral variability is mapped into $Z_1(x)$ and $Z_2(x)$ (Van Der Meer and De Jong, 2000).

3.3.5 Pixel purity index

The pixel purity index (PPI) is used to serve as an automated procedure for separating spectrally purer from more mixed pixels in a remote sensing image. The PPI uses the MNF-transformed results $Z_i(x)$ and reduces the number of pixels to be analyzed for the selection of endmembers

in a remote sensing image by separating pure from mixed pixels. It is computed by repeatedly projecting n-dimensional scatter plots of the pixels on a random unit vector. The extreme pixels, often less than 1% of the total number of pixels fall onto the ends of this unit vector. An image is created using the PPI in which the digital number (DN) of each pixel corresponds to the number of times that pixel was recorded as extreme (Boardman et al., 1995).

3.3.6 The $p - D$ visualiser

The $p - D$ visualiser is an interactive tool to select the end members. The $p - D$ visualiser locates, identifies and clusters the purest pixels. A spatial subset of MNF data which contains only the extreme pixels derived from the PPI, is used as an input to the $p - D$ visualiser.

For a theoretical understanding of the $p - D$ visualiser, pixel spectra can be considered as points in an p-dimensional scatter plot, where p is the number of MNF transformed bands. In the scatterplot, the spectrally purest pixels occur in the corners of the data cloud, whereas spectrally mixed pixels occur within the data cloud. Thus the distribution of these points in $p - D$ space can be used to select the spectral end members (Boardman, 1993).

3.3.7 Mapping selectively-logged sites

Forested areas were separated from non-forested areas (e.g. water bodies, Savannah region and deforested areas) by applying the ISODATA unsupervised classification algorithm on the Landsat image. All non-forested areas were clumped together to classify the image into binary forest/non-forest classes. Afterwards, a median filter was applied to reduce the image noise from the forested areas and to remove small forest-gaps. A forest mask was built from the resulting forest/non-forest classified image. The mask was applied to the soil fraction image obtained from spectral unmixing of the Landsat image. Thus the non-forested areas were masked out to consider only the forested pixels for detecting small canopy gaps due to selective logging.

After applying the forest mask, the pixel values of the forest soil fraction image were examined. Since the log-landing locations show higher fraction values in the forest soil fraction image, a cut-off value was determined visually. Pixels with soil fractions larger than the cut-off value were considered as potential log-landings. The forest soil fraction image was then classified into a binary image using the cut-off value. Pixels with soil fraction values larger than the cut-off but located on the forest edges, roads and along the boundaries of water bodies were excluded from the analysis by manually editing the image.

These pixels were excluded because they may show forest degradation signals which are not associated with selective logging and hence they represent artifacts. Such pixels are easy to identify because of their spatial location, distribution and characteristics; e.g. pixels on roads and forest edges show up continuous linear spatial features (Matricardi et al., 2005).

To estimate the area affected by selective logging around each identifed log-landing location, average distance from each log-landing to its nearest log-landing location was calculated. Half of the calculated distance was taken as the harvesting radius around the log-landings. A circular buffer with sides equal to the calculated distance around each log-landing was generated which represents the extent of the area affected by selective logging.

3.3.8 Spatial statistical analysis of selective logging

Point pattern statistics have been intensively used in forestry sciences to analyze the distribution and spatial dependence among trees in a forest ecosystem (Getis and Franklin, 1987; Mateu and Usó, 1998; Stoyan and Penttinen, 2000; Grabarnik and Särkkä, 2009). The analyses have usually been based on single-tree positions considered as points. Such analyses are useful in small scale studies where it is possible to measure the location of each individual tree and its associated characteristics. When the study area covers a large geographical region, however, we have to rely on the data acquired through remote sensing techniques. The scale at which data are represented is given by the spatial resolution of the remote sensing images (Stein et al., 1998). A spatial disturbance in a forest canopy caused by removal of trees is reflected in the satellite imagery as a change in reflectance value of the canopy structure. These small disturbances, also called canopy gaps, can be represented by points owing to their subtle patterns as detected from the Landsat images and the large scale of the study area considered. Canopy gaps resulting from selective logging, thus, form a spatial point pattern.

In statistical terms, spatial point patterns result from a mixture of first-order and second-order effects. The first-order effects describe the number of logged locations per unit area, and the second-order effects describe the dependency relationships between the locations. First-order effects are determined by the intensity, whereas the second-order effects are usually assessed by the nearest-neighbor distribution functions such as the $F-$ and $G-$ functions (Anwar et al., 2011). Theoretical details of the spatial point processes can be found in Illian et al. (2008). Our interest concerns detection of systematics, i.e. regularity or aggregation (clustering) in the distribution of selectively-logged locations, as deviation from the randomness.

Complete spatial randomness (CSR) indicates that (i) the number of logged locations in a region A follows a homogeneous Poisson distribution with mean $\lambda|A|$, where λ is the constant density; (ii) given n locations $x_i, i = 1, \ldots, n$, the x_i are an independent random sample from the uniform distribution on A (Diggle, 2003). In other words, the density of the point pattern does not vary over the bounded region, and there are no interactions among the locations.

Dependency relationships for interactions can be described by nearest neighbor distances defined as the distance from the ith logged location to the nearest other location in A. The empirical cumulative probability distribution function $\hat{G}$ for the nearest neighbor distances summarizes the incident pattern as:

$$\hat{G}(w) = \frac{\sum_{w_i \leq w} 1}{n},$$

where w_i is a nearest neighbour distance for the ith location. The observed pattern is part of a larger region, where the distribution of selective logging is unknown. Interaction between logged locations lying inside and outside the study region have to be properly accounted for to avoid the edge effects.

A visual inspection of the $\hat{G}(w)$ may reveal the spatial distribution of the logged location. Excess of nearest neighbors at short distances indicates clustering, whereas excess of long distance neighbors refers to regularity.

3.4 Results

Approximately 42% of the total number of pixels in the Landsat image constitutes the artificial zero-reflectance background area in all bands. The image bands were transformed into a new set of bands using the MNF transform. The contribution of each MNF component to the overall information is measured by its eigenvalue. A larger eigenvalue indicates that a component contains more information from the data set. The effect of a large number of the background pixels on the variance-covariance structure of the image is reflected in table 3.1 by the difference in eigenvalues of the MNF transformed bands of the full image as compared to the cropped and rotated image.

3.4.1 Cropped image

Figure 3.1 shows the cropped image constituting the rectangular area inscribed by the image boundary.

Table 3.1: Eigenvalues of the MNF-transformed bands for the full (figure 1.1), cropped (figure 3.1) and rotated (figure 3.6) images

MNF-transformed	Full image	Cropped image	Rotated image
1	419	27	25
2	24	5	5
3	5	4	4
4	3	2	2
5	2	2	2
6	1	1	1

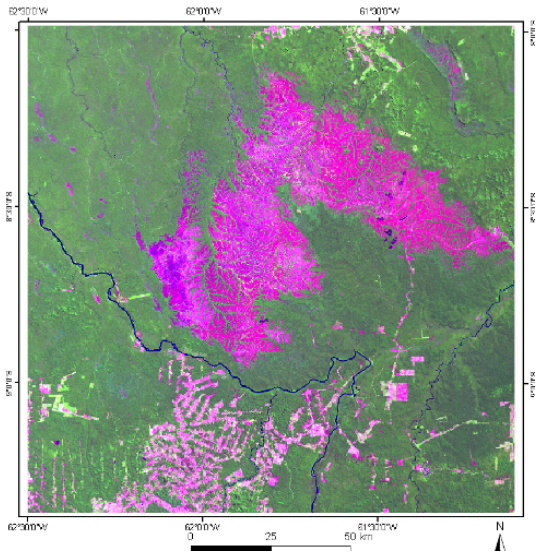


Figure 3.1: Cropped image showing the area inscribed by the original Landsat-5 TM image in R,G,B = 5,4,3

3.4.1.1 Spectral unmixing of the cropped image

Figure 3.2 shows plotted mean spectra of the selected endmember classes for the cropped image. The endmember *Veg* represents green vegetation which is mostly observable in pastures or low vegetation areas, i.e. the vegetation with no self-shadowing; *So* represents bare soil which is mostly associated with unpaved roads and agricultural lands; the class *Sh* represents water, shadowing along ravines and includes the canopy self-shadowing.

The average residual RMSE for this model was 0.003 with a standard deviation of 0.002, which reflected appropriate selection of endmembers.

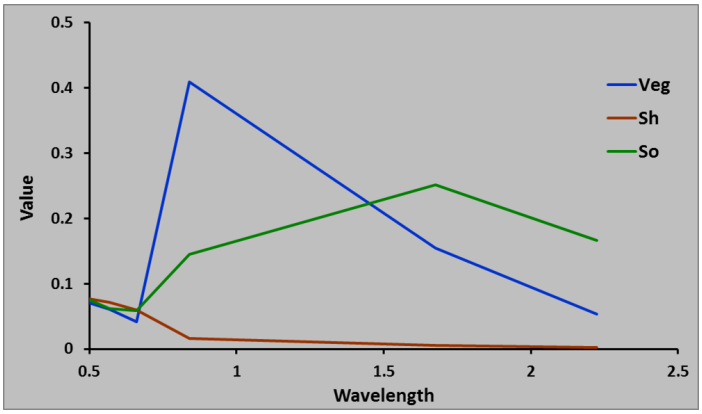


Figure 3.2: Average reflectance spectra of endmember classes: vegetation (Veg), shade (Sh) and soil (So)

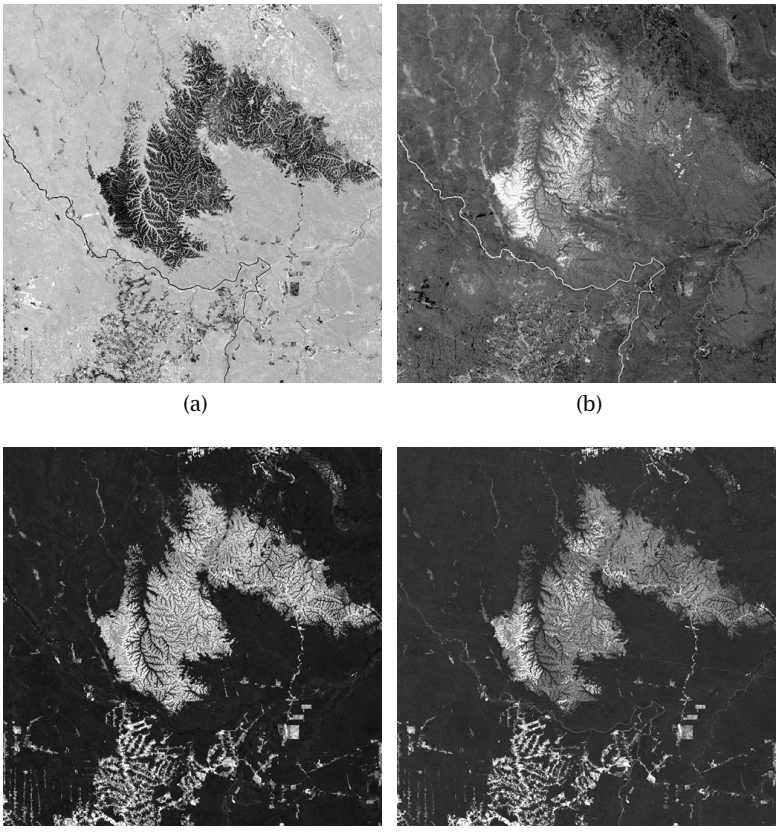
The Veg and Sh fraction images contained at least 99% of all the pixels within 0% to 100%, whereas the So fraction image consisted of at least 90% of the pixels within this range. Figure 3.3 shows the output images for each fraction class and the RMSE image.

On the Veg fraction image, successional forests appear white due to their high Veg fraction values and low shade. Water and bare soils appear black. Pastures and agricultural lands appear in dark grey due to the dominance of soil in such fields. As vegetation grows into mature forest, shade content increases because of self-shadowing of the canopies. This makes the mature forests appear bright grey on the Veg fraction image.

On the Sh fraction image, water appears white, and pastures and agricultural lands appear grey in the Sh fraction image due to their low values of Sh fraction. Mature forests appear bright grey due to the presence of high amounts of canopies self- shadows, whereas successional forests in the north-east of the study area appear dark grey due to their different stand structure. Bare soil shades range between white to dark grey depending on the amount of water content in the soil.

On the So fraction image, roads and bare soil appear white due to their high So fraction values. Pastures and agricultural lands appear white or bright grey, whereas successional and mature forests and water appears dark grey or black due to their very low So fraction values. In selectively logged forests, the log-landings are connected by logging roads which are visible as linear features in the So fraction image.

On the RMSE image, the pixels well-modelled by the selected endmembers model reflect complete noise and hence they appear grey in



(a) (b)

(c) (d)

Figure 3.3: Veg(a), Sh(b), So(c) fraction images and RMSE image(d), derived from the spectral linear unmixing. Bright pixels in the fraction images show a high contribution of the specific endmember material, whereas the dark pixels show lack of the contribution of the endmember in defining the pixel's spectra

the image, whereas the areas not very well modelled appear white. The areas not well-modelled by the endmembers contains mostly soil pixels. This is due to the large variability in the soil types found in the region which a greneral So endmember could not completely capture.

3.4.1.2 Selectively-logged sites detection

After examination of pixel values in the *So* fraction image, a cut-off value of 20% was determined. Pixels with soil fraction 20% or more were considered potential log-landings (Fig. 3.4). Areas of the potential log-landings varied from one to four pixels in the *So* fraction image. The forest soil fraction image was classified into a binary image using the 20% cut-off value. Pixels on the forest edge (30 m area) and at the boundaries of the water bodies were excluded from the analysis by manually removing them. Finally 2195 locations in the image were identified as log-landings.

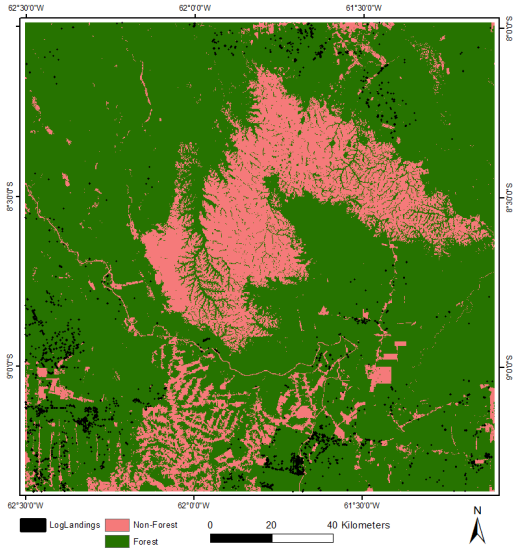


Figure 3.4: Locations of log landings (black dots) detected from the *So* fraction image, and the forest transacts A and B on the forest/non-forest classified map

To estimate the area affected by selective logging around each log-landing location identified in our study area, 50 detected log-landing locations were randomly selected from the image. Selection was made from the parts of the soil fraction image where clearly visible patterns of logging roads and log-landings locations were found. Average distance from each log-landing to its nearest log-landing location was found to be 400 m. Half of the average distance was taken as the harvesting radius. A circular buffer with radius of 200 m was, therefore, created to represent the extent of area affected by selective logging around each

log-landing location. Two forest transects A and B, where patterns of mechanized logging were visible, were chosen for detailed display of results (Fig. 3.5).

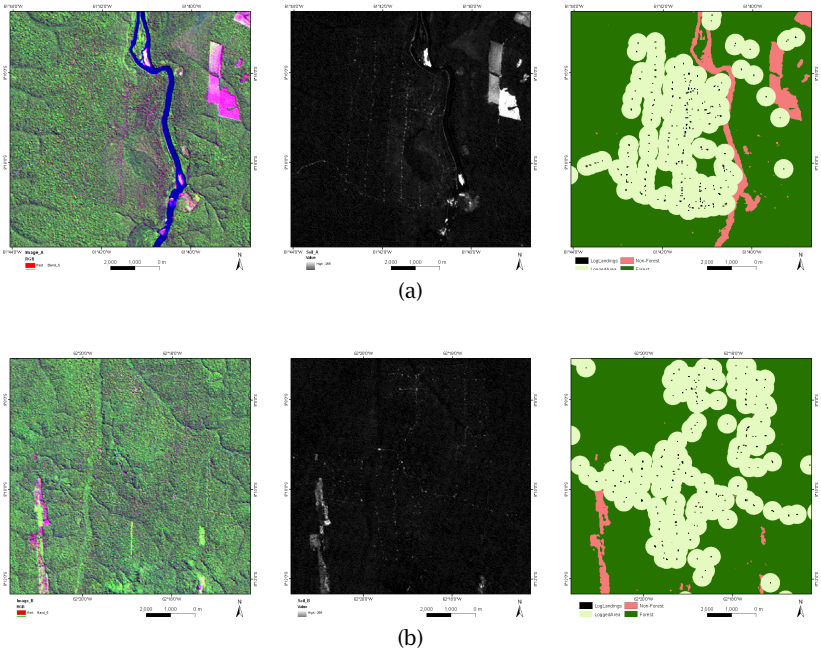


Figure 3.5: Original image, *So* fraction image and logged areas for forest transact-A (a) and transact-B (b). Selectively logged areas are recognized from the *So* fraction image by the linear features formed by log landings and logged roads

3.4.2 Validation of results

Pixel values of each CBERS-2B image downloaded for the study area, were visually inspected and a minimum threshold value was determined for each image to represent the degradation signals. Binary masks were then applied to the images using the threshold values. After masking, a median filter was applied to remove the image speckle caused by illumination variations of the canopy. The procedure highlighted the locations of forest degradation in the CBERS images. Afterwards, a total of 50 log-landings were randomly selected from the parts of the Landsat image where the CBERS data were available for validation. All the log-

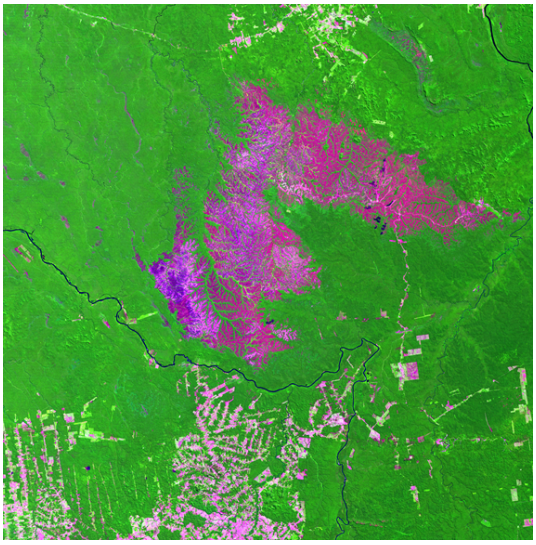


Figure 3.6: Image (R,G,B = 5,4,3) obtained by rotating the original Landsat-5 TM image to 8.0^o anti-clockwise

landings locations, with the exception of six locations, could be located in the CBERS images. Hence, the overall accuracy of the log-landings detection procedure was found to be 88%.

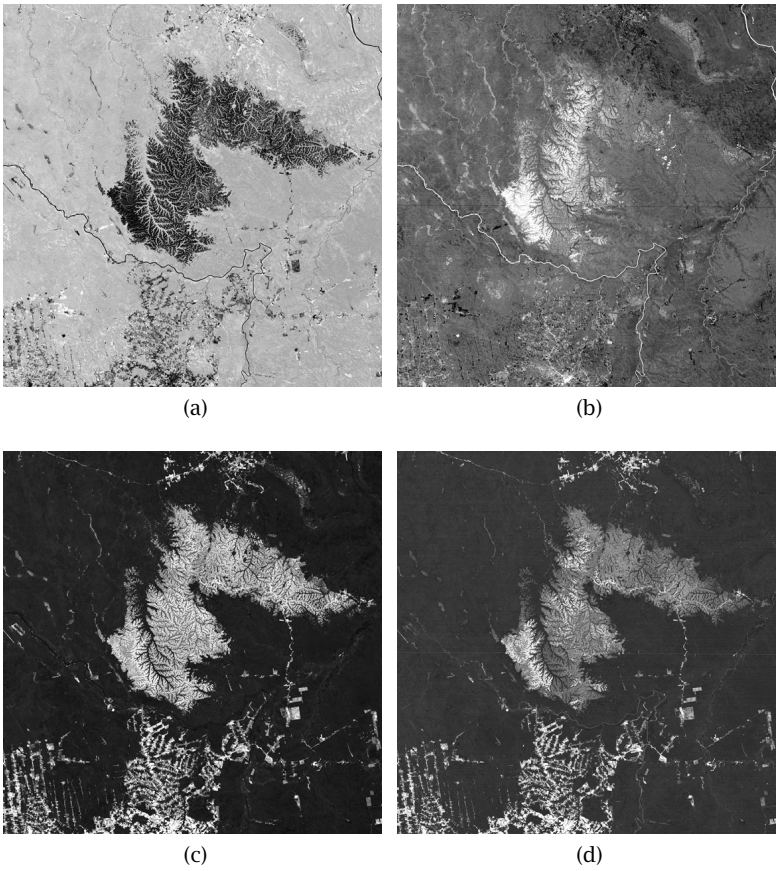
3.4.3 Rotated image

Figure 3.6 shows the Landsat image rotated to 8.2^o in the anti-clockwise direction and then cropped to exclude the artificial zero-reflectance background area.

The pixels used as endmembers for unmixing of the cropped image were located in the rotated image. The same pixels were used as input image-endmembers to the unmixing of the rotated image. Figure 3.7shows the output images for each fraction class of the rotated image and the RMSE image.

The average residual RMSE for this model was 0.003 with a standard deviation of 0.002. The Veg and Sh fraction images contained at least 99% of all the pixels within 0% to 100%, whereas the So fraction image consisted of at least 88% of the pixels within this range.

The methodology described in section 3.4.1.2 was repeated for the rotated image and the area affected by selective logging was detected. 3276 locations in the image were identified as log-landings. The image was re-rotated back to its original orientation (Fig. 3.8).



(a) (b)

(c) (d)

Figure 3.7: Veg(a), Sh(b), So(c) fraction images and $RMSE$ image(d), derived from the spectral linear unmixing of the rotated image. Bright pixels in the fraction images show a high contribution of the specific endmember material, whereas the dark pixels show lack of the contribution of the endmember in defining the pixel's spectra

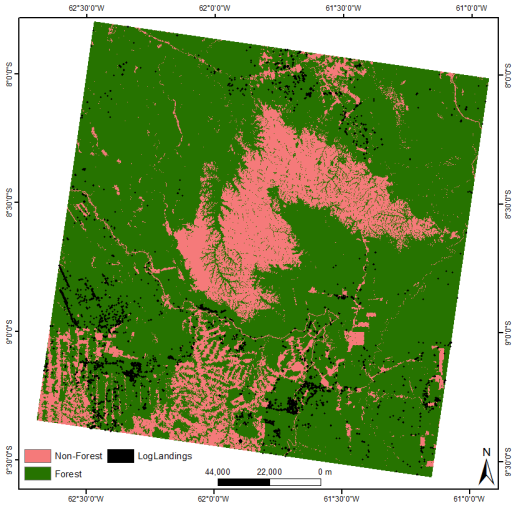


Figure 3.8: Locations of loglandings (black dots) detected from the *So* fraction image, and the forest transacts A and B on the forest/non-forest classified map

3.4.4 Comparison

The results obtained for the cropped and rotated images were compared in terms of the total number of detected log-landings and their locations in the area shared by both images. The comparison revealed good match between the results, with 2195 log-landings in the area shared by both images. The rotated image covers a larger area than the cropped image and hence we can detect the selective logging over a wider area. The pixels in the re-rotated image were, however, shifted up to approximately 30 m (1 pixel). Shifting in the pixels locations caused the positional error due to the nearest-neighbour resampling technique. The nearest-neighbour resampling technique preserves the original pixels values but shifts the pixels from their original locations. The magnitude of shifting depends upon the angle of rotation. For the Landsat image used in the study, the orientation angle equals 8.2^o which caused a positional error of approximately 30 m in the pixel location.

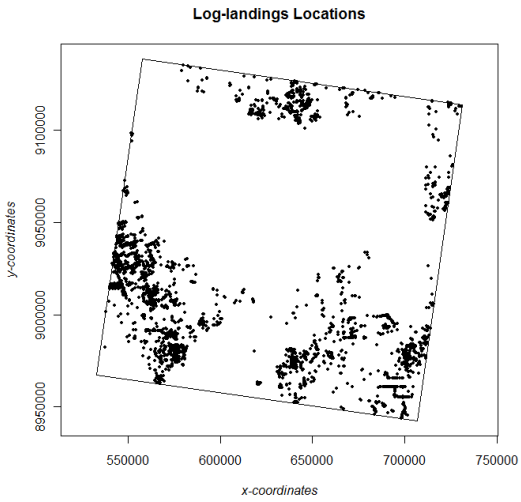


Figure 3.9: Point pattern constituting the log-landings locations in the study area

3.4.5 Clustering

Centroids of the selectively-logged locations (Fig. 3.8) were taken as points for spatial point pattern analysis of selective logging. Figure 3.9 shows the point pattern thus formed of the selectively-logged locations.

To estimate the first-order effects of the point pattern of selective logging, logging intensity was estimated and plotted using kernel smoothing (Baddeley, 2008) (Fig. 3.10). Issues involved in suitable selection of bandwidth for kernel smoothing are discussed in detail in (Illian et al., 2008). After trying a number of values, bandwidth was finally selected to be 20000 as it gave an appropriate generalisation of the variability in intensity of selective logging occurrences. The intensity is heterogeneous throughout the study area as figure 3.10 shows. A high concentration of selective logging is found in the southwestern part of the study area, making it a hot-spot for logging operations.

The second-order effects of the point pattern formed by selectively-logged locations were investigated using distance-based $G-$ function. Figure 3.11 shows the estimated and the theoretical $G-$ functions for selectively-logged locations. The estimated curves lie above the theoretical curve of the Poisson process, which is an evidence that the selective-logging pattern exhibits clustering. A close inspection of the curves, however, reveals that the curves for estimated $G-$ function

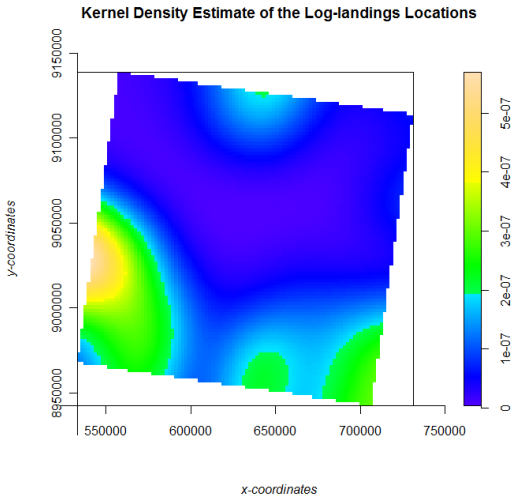


Figure 3.10: Density of the log-landings locations throughout the study area. The south-western part shows the highest concentration of log-landings in the area

matches the curve of the nearest- neighbour $G-$ function of a Poisson process upto a distance of about 50 m. This means that the log-landings locations exhibit randomness at short distances upto 50 m, followed by strong clustering.

3.5 Discussion

Real-time monitoring of selective logging over the vast Amazonian forests requires observations over large areas. The medium-resolution Landsat images offer the advantage of large area coverage with spatial resolution enough for detecting degraded forests. Detection of selective logging using the spectral unmixing technique is hindered by the artificial background introduced due to the orientation of the Landsat images. The artificial background consists of only zero-reflectance pixels in all bands, which affect the variance-covariance structure of the image bands. A large number of the zero-reflectance pixels in all bands equalises bands variances and gives false indication of almost perfect between-band correlation. In such case, the MNF transformation fails to provide realistic information about the bands contents and only the first MNF band shows significant amount of information as compared to the

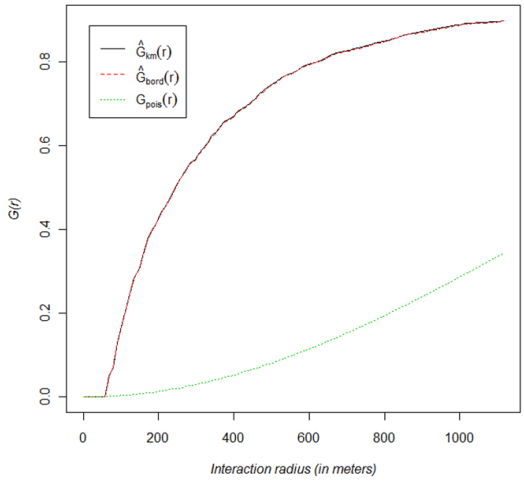


Figure 3.11: The nearest-neighbour $G-$function calculated for point pattern comprising of the log-landings locations

other bands. The amount of information reduction in successive bands is manifested by the difference between the eigenvalues of the first and successive MNF components (table 1). Furthermore, the zero-reflectance background pixels are identified as the purest pixels in the image by the PPI, ignoring the true image data. Since the pixels in the true image are not accounted for by the automated procedure of endmember detection, the spectral unmixing technique can not be applied to the geometrically corrected Landsat images.

Image cropping and image rotation provide quick and easy solutions to the problem discussed above. Cropping, however, results in smaller image extent, whereas rotation introduces small (upto 30 m) positional errors in the image. Image rotation may serve as a better solution, since a small positional error upto 30 m is quite negligible owing to the total area coverage of the Landsat scene which is approximately 30000 km^2. This approach may help to delineate the regions where selective logging operations are concentrated and hence to direct the forest conservation efforts towards those regions.

An effective monitoring of selective logging in Amazonia requires a better understanding of the historic patterns of selective logging. Taking into account the historic patterns of forest degradation, however, is hindered by unavailability of the field data. The methodology by Souza and Barreto (2000), used in this study provides an effective strategy to overcome the problem. Moreover, the more recent techniques such as

those provided by Souza Jr. et al. (2005) and Broadbent et al. (2008), employed for selective logging sites detection, rely heavily on the field data and their methodologies are computationally extensive and complex. Hence they do not offer an attractive solution for real-time monitoring of degradation over vast regions of Amazonian forests. The methodology used in this study is relatively simple and easy to apply and, although may be less accurate, can at least pin-point the areas under most immediate threat of degradation by selective logging.

For spatial point pattern analysis of the selectively-logged locations, the canopy gaps centroids were taken as points. Such approximation may lead to valid results if the sizes of the objects are small in comparison to the spatial scales investigated (Nuske et al., 2009). However, the second-order statistics which involve the nearest-neighbour distances between the objects must be interpreted with caution. In that case, the areas of the spatial objects need to be considered. The distances between their centroids must be replaced by the distances between their respective boundaries to obtain realistic results and to avoid false indications of hard-core distances between the points. Fitting a spatial point pattern model, which is suitable to explain the clustered pattern of log-landings locations, may be a further research direction in order to assess the influence of geographic and socio-economic factors on the distribution of selective logging in the study region.

3.6 Conclusions

This study leads to the following conclusions that are relevant for studies on deforestation in Amazonian forest with a selective logging detection.

1. The endmembers used in this study are well suited to identify selective logging locations. They represent basic components of a degraded forest environment.
2. For the currently available remote sensing images with a resolution of 30 m, spectral unmixing is a well-suited technique to identify patterns of selective logging locations. It may either avoid the collection of field data or help to reduce sampling efforts.
3. The image rotation has been suggested to be a better approach since it offers the advantage of selective logging detection over a larger area. Positional errors up to 30 m, however, have to be compromised.
4. A spatial statistical analysis of the detected locations shows a strong clustering within the study area. This is an evidence that deforestation starts at a limited number of locations and spreads from there.

In all, we conclude that a proper and dedicated use of remote sensing analysis and spatial statistics is a promising way ahead for better understanding and possibly reducing deforestation.

4 Spatial analysis of selective logging to investigate early forest LULC change

Abstract

Selective logging gives currently a major contribution to ongoing deforestation in the Brazilian Amazonia. The spatial distribution of log landing sites (LLS), i.e. the sites where logged trees are collected, serves as a proxy to the intensity of selective logging activities. In this study we analyzed the LLS pattern in a study area that has a rapid deforestation. Actual LLS locations were extracted from a Landsat image of 2000 that covers a large part of the study area. We first used the inhomogeneous J-function. A kernel bandwidth of 20 km best modelled the non-stationarity, showing a strongly clustered LLS distribution. Second, the Area-interaction point process model incorporating information about distance of LLS to roads and to clear-cut deforested areas was applied. The model well explained the clustered LLS pattern and showed a significant effect of distance to roads. We concluded that this spatial statistical study helped to quantify and better understand the LLS pattern.

4.1 Introduction

Landscape spatial patterns refer to the spatial heterogeneity in a landscape. Spatial heterogeneity means the composition, configuration and, in a broader sense, even the temporal aspects of the heterogeneity (Gustafson, 1998). Spatial patterns reflect processes which have been operating in the past and may be present in the future (Law et al., 2009).

[1] This chapter is based on the following paper
Anwar, S., Stein, A., 2014. Use of spatial statistics to investigate early forest degradation activities as detected from satellite images. *Submitted to Spatial Statistics*

In a forest environment, spatial patterns may result from a number of different ecological processes, examples of which may include growth and mortality, interaction of silvicultural management, land use and climate (Comas and Mateu, 2007). As there is a strong link between patterns and processes (Gustafson, 1998), ecologists study spatial patterns to infer the existence of underlying ecological processes related to the phenomena under study (Watt, 1947). Knowledge of spatial patterns in a forest environment, therefore, may assist in management of forest resources.

In the Brazilian Amazonian forests, selective logging for timber is a major source of forest degradation (Uhl et al., 1991; Gerwing, 2002). Adverse effects of selective logging on the Amazonian forests include damages to forest phenology (Koltunov et al., 2009), increasing vulnerability of a forest to fire (Gerwing, 2002; Fearnside, 2005), forest fragmentation (Broadbent et al., 2008) and wide spread of deforestation (Uhl et al., 1991; Fearnside, 2005). An important challenge faced by forest researchers today is to detect and analyse the locations and patterns of forest perforations caused by selective logging for timber. Selective logging operations are almost impossible to monitor, as much of the Amazonian region is inaccessible. Remote sensing, due to its synoptic view and fast coverage, may serve as a viable source for the purpose. Log-landing sites(LLS), the locations where the collected timber is stored, can be detected from remote sensing images and may serve as a proxy for the selective logging activities in a surrounding area (Anwar and Stein, 2012).

There has been intensive work on analysing the deforestation patterns and processes in the Brazilian Amazonia and efforts have been made to model the deforestation processes using different spatial analytical tools (Apan and A., 1998; Alvis et al., 1999). Spatially explicit analysis and modelling of the locations of forest degradation is, however, lacking. Due to small spatial extent of the LLS, it is natural to represent them as points in the maps derived from remote sensing images. A spatial point pattern analysis, then, may serve as an appropriate tool for analysing the process that determines the LLS distribution. On the other hand, as logging operations also vary in time (Matricardi et al., 2005); a spatial-temporal statistical analysis of the spatial distribution of selective logging is necessary to reveal its important temporal characteristics (Anwar and Stein, 2014a) .

A spatial point pattern is usually a single realization of a spatial point process. The assumption of stationarity is often made to reduce the parameter space and to allow parameter estimation (Cressie, 1993) of a spatial point process. By stationarity it is understood that all properties of the spatial process are invariant under translation. We thus assume that the LLS density is constant across space. Real forest configurations,

however, are seldom stationary. Inhomogeneity may arise as a result of the inherent spatial variability caused by environmental factors, or due to the interactions between the locations of the phenomenon under study or both. In spatial point process theory, the term 'interaction' refers to the probability of points of the same or different types occurring in close proximity (Illian and Burslem, 2007). In a forest structure, different factors such as environmental heterogeneity and geographical configurations (e.g. mountain slope) can cause inhomogeneity (Comas et al., 2009). For a LLS pattern, the stationarity assumption could be violated if there are large areas of low or zero density of LLS over the region under study.

Dealing with inhomogeneity has been a focus of spatial statistical research (Law et al., 2009). The efforts have been directed towards finding appropriate tools for the analysis and formulating suitable models to address nonstationarity of the point processes. Baddeley et al. (2000) developed an inhomogeneous version of Ripley's K-function (Ripley, 1977) as an important contribution for investigating the intensity structures and analysing the spatial distribution of point patterns. Another summary measure, the inhomogeneous J-function (J_{inhom}) was proposed by Van Lieshout (2011). The J_{inhom} is receiving popularity because, unlike the inhomogeneous K-function, its definition does not depend upon the choice of origin and hence is computationally convenient. A tricky issue involved in calculation of inhomogeneous summary functions is the choice of kernel bandwidth (Law et al., 2009), since there is no rigorous mathematical theory available for its optimal choice (Illian et al., 2008).

Modelling of a nonstationary process is done using Monte Carlo Markov Chain (MCMC) methods. In the presence of interaction among LLS, such methods may become computationally extensive as they need to incorporate interaction terms as well as the spatially varying LLS density (Baddeley et al., 2000). The LLS data are commonly found to be of an inhomogeneous structure and a number of different ecological and geographic factors may influence their distribution pattern. A model can be considered realistic only if it is capable of modelling the inhomogeneity by incorporating available information on the environmental and geographic variables as covariates, as well as the interaction dependent upon locations.

The present study focuses on the assessment of LLS patterns. The objective is to discover important spatial characteristics of selective logging in the south-western part of the Brazilian Amazonia using J_{inhom} and then modelling a LLS pattern using a model suitable for inhomogeneous processes. While the purely spatial aspects are dealt with in this paper, the spatial-temporal aspects are treated in (Anwar and Stein, 2014a).

4.2 Study area and data description

The study area is described in detail in Chapter 1. A Landsat 5-TM image of the study area for year 2000, with row 231 and path 66, was downloaded from the United States Geological Survey (USGS) website (http://earthexplorer.usgs.gov/). The images on the website are processed to Standard Terrain Correction (Level 1T), which means that the images are radiometrically and geometrically corrected.

Roads were digitized manually from the soil faction image derived from the linear spectral unmixing of the Landsat image at scale 1: 50,000. The roads show a higher fraction of soil and clearly linear pattern in the soil fraction image. A pixel image of the study area with pixel values representing the distance of each pixel from the closest road segment was used as a covariate information (Fig. 4.1(a)) in modelling.

Supervised classfication of the Landsat image was performed and different types of landcover were aggregated to classify the image into three classes: forest, nonforest (constituting mostly of the Savanna region in the upper middle part of the image) and clearcut deforested areas. Boundaries of the clear-cut deforested polygons were vectorized and a pixel image of the study area was created representing the distance of each pixel from the nearest deforested polygon (Fig. 4.1(b)). This image was also used as covariate information for describing the LLS distribution at the modelling stage. This was done to assess the influence, if any, of the proximity to deforested areas on the LLS distribution.

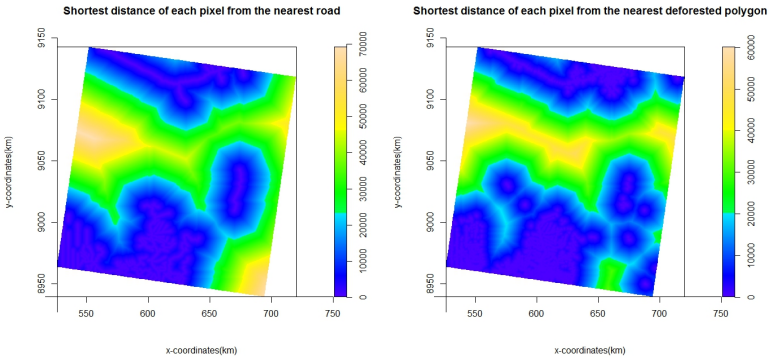
The LLS were detected from the Landsat images by applying the methodology described in Anwar and Stein (2012) which is based on the linear spectral unmixing of the images. Three fraction images were calculated representing green vegetation, shade and soil. The soil fraction image was used to detect the LLS within the forested areas in the Landsat image.

4.3 Methodology

4.3.1 Point pattern of the LLS

At the scale of the area covered by a Landsat image, the LLS can be considered as a spatial point pattern. These patterns can be summarized by a mixture of first- and second-order effects. First-order effects are described by the intensity $\lambda(x)$ i.e. the expected number of LLS per unit area. Mathematically,

$$\lambda(x) = \lim_{|dx| \to 0} \{E\left[N(d(x))\right] / |dx|\}$$



(a) (b)

Figure 4.1: Distance (in meters) of each pixel to (a) the nearest road segment and (b) the nearest clear-cut deforested polygon

where dx is a small region at site x of the LLS pattern L, $|dx|$ being its area and $N(dx)$ is the number of LLS in dx.

Second-order effects are given by inter-LLS interactions $\lambda_2(x_i, x_j)$ that describe the stochastic dependence between two LLS x_i and x_j. They refer to the variability in the number of LLS per unit area for the whole spatial pattern of these sites. Mathematically,

$$\lambda_2(x_i, x_j) = \lim_{|dx_i|,|dx_j|\to 0} \left\{ E\left[N(d(x_i))N(d(x_j)) \right] / |dx_i||dx_j| \right\}.$$

Basic concepts and analytical methods are in e.g. Diggle (2003).

4.3.2 Homogeneous and inhomogeneous point processes

A spatial point pattern is usually a single realization of a spatial point process. The assumption of stationarity is, therefore, often made to reduce the parameter space and to allow parameter estimation (Cressie, 1993). Stationarity for a LLS pattern refers to spatial homogeneity which means that all the properties of the spatial process resulting in the observed LLS pattern do not vary across space (Illian et al., 2008). More specifically, a LLS generating process $L = x_n$ is stationary if its distribution is invariant under translation. Thus the translated process $L_x = x_n + a$ has the same distribution for all translation vectors a in R^d. In that case, the first-order summary characteristic given by the density of a homogeneous LLS pattern does not depend upon the sites x_i. The density function $\lambda(x)$ can then be replaced by the intensity λ,

which is the spatial average density of LLS and is estimated by dividing the number of LLS by the area $|A|$ of the region A. For further details about stationary point processes we refer to (Illian et al., 2008).

The assumption of stationarity is an important limitation of the methods of spatial point processes (Law et al., 2009). For a LLS pattern, the stationarity assumption could be violated due to variations in the existance of social, economic and ecological factors responsible for instigating and accelerating selective logging, and hence to an LLS of increased intensity, as well as to spatial interactions. In spatial point process theory, the term interaction refers to the probability of LLS showing a high density locally (Illian and Burslem, 2007). Interaction between the LLS may be described by the fact that new forested areas are selectively logged in continuation to the previously logged forests. Thus selective logging in one part of the forest encourages further selective logging in its proximity, thereby causing interaction between the LLS. For these reasons, it is implausible to expect stationarity in LLS.

4.3.2.1 Inhomogeneous J-function

Extension of the J-function to inhomogeneous spatial point configurations has been proposed by Van Lieshout (2011). The inhomogeneous J-function (J_{inhom}) compensates for the inhomogeneity by weighting the stationary L process by the intensity. Mathematically, it is expressed as:

$$J_{inhom}(r) = \frac{G^{!a}(1 - u_r^a)}{G(1 - u_r^a)}, \tag{4.1}$$

where

$$u_r^a(x) = \frac{\overline{\lambda} B_x(a, r)}{\lambda(x)}, \tag{4.2}$$

for $r \geq 0$ and $a \in \mathbf{R}^d$. In the above equations, $B_x(a, r)$ represents a circle of radius r centred at a, $\lambda(x)$ is the intensity at location x, $\overline{\lambda} = inf_x \lambda(x)$ which is bounded away from zero, G is the generating functional of the Palm distribution P, and $G^{!a}$ is the generating functional of the reduced Palm distribution $P^{!a}$. The generating functional is defined as the extension of the multivariate probability generating function (p.g.f.) to the generating function of an infinite set of random variables (Westcott, 1972).

For the intensity-reweighted moment stationary process L,

$$F_{inhom}(r) = 1 - G(1 - u_r^a(x))$$

and

$$G_{inhom}(r) = 1 - G^{!a}(1 - u_r^a(x)).$$

Then, by definition:

$$J_{inhom}(r) = \frac{1 - G_{inhom}(r)}{1 - F_{inhom}(r)} \tag{4.3}$$

The J_{inhom} offers a suitable choice to capture the inhomogeneity in the LLS pattern.

4.3.2.2 Estimation of λ and choice of kernel bandwidth

The intensity function $\lambda(x)$ in Eq. 4.2 is assumed to be bounded away from zero to avoid the denominator of the equation being zero. The intensity is generally unknown and estimated nonparametrically by the techniques such as the kernel smoothing. A kernel estimator of intensity takes the form:

$$\hat{\lambda}_h(x) = \sum_{i=1}^{n} \kappa_h(x - x_i), \tag{4.4}$$

where $\kappa_h(.)$ is a bivariate probability density function symmetric around the origin, for example $\kappa_h(u) = h^{-2}\kappa(u/h)$ and h, the bandwidth, is a positive real number.

The main challenge in estimation of the intensity function is the suitable choice of kernel bandwidth h. A small value of h may result in a too spiky density surface, whereas a large h leads to smoother surfaces that may ignore local features of the density of the point pattern (Law et al., 2009). In literature, no optimal choice of h is provided. In our study we used a trial and error procedure. We considered values of the bandwidth equal to 10, 20, 30, 40 and 50, and plotted the estimated intensity surface for each of these values. The results were compared to select an optimal value that minimized the difference between the estimated density and the true density.

4.3.3 Inhomogeneous Markov point process model

It is of interest to determine the process which generated the inhomogeneous distribution of the LLS data. To reflect the ground situation, we need a mathematical model that describes the inhomogeneity by accounting for the spatial variability caused by covariates effects as well as the interaction between the LLS. Markov point processes, or more generally the Gibbs point processes, are the natural choice to take care of the inter-LLS interactions.

4.3.3.1 Pairwise-interaction point processes

The Pairwise-interaction point processes are the simplest and most commonly used subclass of the Gibbs processes in which the interactions

occur only between pairs of points. For the pairwise interaction processes, the conditional probability of finding a log-landing at location u given complete realization of L is given by the conditional intensity function of the form:

$$\lambda(u, L) = b(u) \prod_{i=1}^{n(L)} c(u, x_i)$$

where $b(u)$ is the first-order term and $c(u, v)$ is the second-order or pairwise interaction term which measures dependence between LLS. The pairwise interaction processes model the interaction structure of point patterns. These processes are good models for repulsive (regular) patterns but they do not describe clustered patterns in a sufficient way (Law et al., 2009). An example is the Strauss model which was originally developed to model clustered patterns but later turned out to be suitable only for regular patterns. The density function of this model proved to be non-integrable for the parameter values corresponding to the desired clustering.

4.3.3.2 Area-interaction point process

To avoid the limitations of pairwise point processes, a higher-order interaction model, named Area-interaction point process model, has been developed. The Area-interaction model is also a Markov point process with interactions of infinite order. The conditional intensity function of the Area-interaction process is given by the equation

$$\lambda(u, L) = \beta \gamma^{-B(u,L)}$$

where $B(u, L) = A(L \cup u) - A(L)$ is the area of that part of the disc of radius r centred on u that is not covered by discs of radius r centred at the other $x_i \in L$. As each LLS represents the area logged within a circle of radius r, we may interpret $B(u, L)$ as the area which has not been logged yet and where a new LLS point may occur. A value $\gamma > 1$ gives a higher chance for the specification of such a new loglanding site, whereas for $\gamma < 1$ a new loglanding site is less likely to emerge. Thus, $\gamma = 1$ corresponds to a Poisson process, $\gamma < 1$ produces a regular process and $\gamma > 1$ a clustered process.

4.3.4 Model diagnostics

Various formal and informal methods to check the goodness-of-fit of a fitted point process model are available (Baddeley, 2008). Formal methods include χ^2 goodness-of-fit tests and Monte Carlo tests, whereas the informal methods include the residual analysis of a fitted point process

model (Diggle, 2003). For Gibbs models χ^2 goodness-of-fit test and Monte Carlo tests are theoretically not supported. Instead, the common K- and G-functions are used to simulate the critical envelopes for the fitted models which are then used as tools for checking the validation of the fitted Gibbs models (Anwar et al., 2011). Appropriateness of the inter-LLS interaction function in fitted Gibbs model is determined using an 'informal' validation tool known as QQ-plot of the residuals (Baddeley, 2010a). It compares empirical quantiles of the smoothed residual field to the corresponding expected empirical quantiles under the fitted model (Baddeley et al., 2005).

4.4 Results

4.4.1 LLS point pattern and J_{inhom} function

From the Landsat image of year 2000, 650 LLS were detected and plotted in Fig. 4.2.

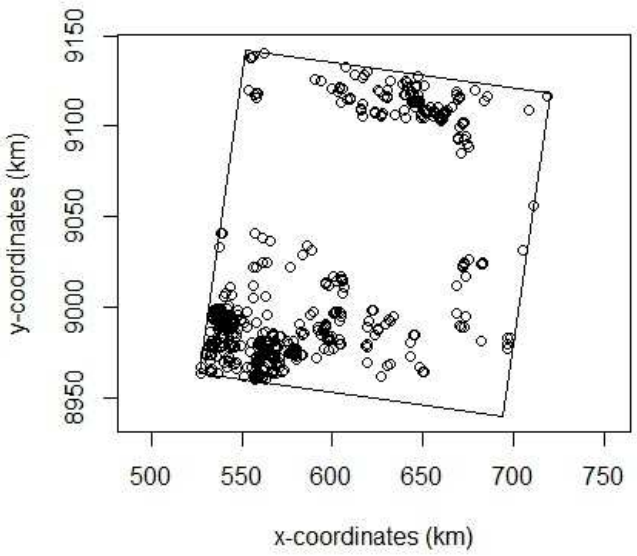
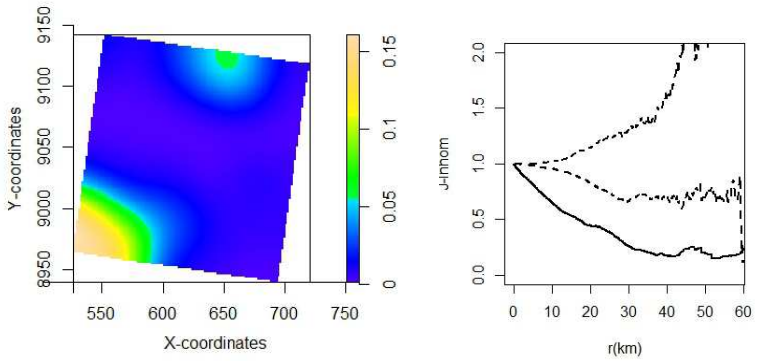


Figure 4.2: LLS point pattern in the year 2000, showing a large cluster of LLS in the lower left corner of the study area and another cluster in the upper middle part of the area. The central part of the image does not indicate any logging activity mostly due to the presence of Savanna region in that area



(a) (b)

Figure 4.3: (a) kernel density estimate of intensity with kernel size = 20 km and (b) Inhomogeneous J-function with this intensity

The figure clearly reflects inhomogeneity in LLS pattern. There are large areas within the study region with almost no LLS, whereas in other parts there exist large clusters. Hence the LLS apparently exhibit a inhomogeneous and clustered pattern.

Intensity of the LLS was estimated non-parametrically using kernel density estimator. The density surfaces for kernel sizes equal to 10, 20, 30, 40 and 50 km were compared with the LLS pattern shown in Fig. 4.2. The kernel size equal to 20 km was found to be a suitable choice to represent the variability in the LLS density. Fig. 4.3a shows the intensity surface calculated using the kernel density estimator with kernel size equal to 20 km. The same density estimate was used to calculate J_{inhom} plotted in Fig. 4.3b. We also calculated the upper and lower envelopes from 19 simulations of a Poisson process in the sampling window to quantify the deviation of the observed pattern from a random process. We found that the observed J_{inhom} function lies below the simulated envelopes of the theoretical J_{inhom} function, which is an evidence of a clustered LLS pattern. The result of the calculated function agrees with the visual assessment of clustering of the LLS pattern and hence the choice of the kernel bandwidth appears reliable.

Figure 4.4 shows the density and the J_{inhom} functions for the LLS pattern for the different values of the kernel bandwidth. Figures **??**(a) shows the intensity surface and the inhomogeneous J-function calculated using the kernel density estimate with kernel bandwidth equal to 10 km. The density surface is tightly concentrated around the observed LLS and therefore seems too spiky. This inappropriate choice of the

kernel bandwidth is reflected in the calculated J_{inhom} which shows a completely random LLS pattern which does not reflect the ground pattern in Fig. 4.2. On the other hand, as can be observed from Fig. 4.4, the values of the kernel bandwidth larger than 20 km smooth out the density surface up to the extent that important local variations in the LLS pattern are ignored. Effect of the bandwidth larger than the suitable choice can be observed in the J_{inhom} functions plotted in Figure 4.4.

Ideally the range of interaction distances r should represent the actual distances between the LLS. A larger value of the kernel bandwidth spreads the influence of a LLS over a greater distance, thereby reducing the interaction distances between the LLS. This causes the smaller range of interaction distance over which the J_{inhom} function is calculated. Here, by 'influence' of a LLS we mean the area logged around a LLS. Beyond the range of spatial interaction between LLS, the J-function becomes almost constant. From the figure we observe that as the kernel size increases, the effective range of interaction distances decreases.

From the calculated J_{inhom} functions we observe that as the value of r increases beyond its effective range, the simulated envelopes span over a wider range and the relative noises in the simulated envelopes also increase. This is due to the fact that interaction between the LLS decreases with increasing interaction distances. Beyond the effective range of interaction distances, the pattern becomes increasingly random, hence causing the wider ranges of simulated envelopes. Also the relative noise in the simulated envelopes of the J-function increases beyond the effective range of interaction distances.

4.4.2 Spatial modelling

To account for the inhomogeneity in the intensity of LLS pattern, different forms of intensity functions for the Area-interaction model were fitted including linear, quadratic and cubic trends, using different values of the interaction radius r within the interval 1-5. Performance of the fitted models was tested by comparing their AIC values. The intensity function which is a linear function of the spatial locations was found to be suitable to represent the LLS distribution because the higher-order polynomials did not result in significant improvement in the AIC values.

Next, to estimate the optimal value of r for the Area-interaction process, we first investigated the residuals from the Area-interaction model with log-linear form of the intensity function, fitted to the LLS data for all different values of r within the interval 1-5. We investigated the residuals from the fitted models using the famous QQ-plot which is a commonly used plot for Gibbs models. We found that the QQ-plot for $r = 3$ was showing a better fit for the Area-interaction model, although still not the perfect one.

The influence of geographic information on the distribution of LLS was investigated by incorporating them as covariate information into the fitted model. The available covariate information included the location of roads and the clear-cut deforested areas in the study region (Fig. 4.1). The performance of the fitted intensity functions was quantified as AIC values. QQ-plots were also drawn to help assess the improvement in the performance of the fitted models, following Baddeley (2010a). Fig. 4.5 shows trends fitted by the three models and the QQ-plots for each model, alongwith their log-intensity functions and the respective AIC values.

Comparing the trends fitted by the intensity functions of Model 1 and Model 2 and their respective QQ-plots in Fig. 4.5, we observe that with inclusion of covariate information on the presence of roads in the study area, the model fitting results are much improved. However, comparing Model 2 and Model 3, we find that inclusion of covariate information on presence of clear-cut deforested areas in the study area does not add any further to model improvement. We also find that the AIC value of the the Area-interaction model remains unaffected when we incorporate the covariate information on the deforested polygons in addition to the covariate information on the presence of roads in the area. This means that the information on the locations of deforested areas does not serve as an informative covariate when the information on presence of roads has already been incorporated in the model. Furthermore, comparing the patterns found in Fig. 4.1a and 4.1b, we observe that the deforested areas are concentrated along the roads within the study region and therefore roads and the deforested areas show a similar distribution pattern. Hence both covariates can serve as proxy to each other. These observations provide an interesting evidence that within the study area, the roads are a major factor determining the distribution of LLS as well as causing clear-cut deforestation of the Amazonian forests.

Table 4.1 shows the estimated values of the parameters of the fitted log-intensity functions formulated in Fig. 4.5 and the interaction parameters of the respective Area-interaction models:

	a	x $\times 10^{-4}$	y $\times 10^{-4}$	DR $\times 10^{-5}$	DC $\times 10^{-5}$	Interaction parameter
Model 1	-7.47	-8.47	1.01			1.25
Model 2	-9.27	-4.23	3.08	-1.69		1.20
Model 3	-8.05	-5.75	1.88	-0.68	-1.76	1.20

Table 4.1: Estimated parameters of the fitted log-intensity functions. The last column shows the values of the estimated interaction parameter

Comparing Model 1 and Model 2 we observe that with inclusion of the covariate information on the location of roads, interaction parameter decreased. This is because a part of the variability in the LLS distribution has now been explained by the covariate information on roads. However when we include the covariate information on the presence of deforested areas in the model, the interaction parameter does not decrease any further which indicates that this additional covariate is not much informative. However by including alternative covariates, the interaction parameter might decrease and would converge to one i.e. the LLS pattern would tend to a non-stationary Poisson process. Thus if all of the variability in the LLS pattern could be explained in terms of covariate effects, the interaction among the LLS would become negligible and the pattern would exhibit characteristics of a non-stationary Poisson process.

Critical envelopes were simulated for Model 2 using the inhomogeneous K-summary function to provide an additional evidence of the suitability of our final fitted model to the LLS data of year 2000 in (Fig. 4.6). The figure shows that the model fits well to the LLS pattern.

4.5 Discussion

This study proposes the application of spatial point process methodology for an improved understanding and analysis of the LLS spatial pattern which serve as proxy for selective logging operation in the study area.

Our study demonstrates that determining the presence of homogeneity in the LLS pattern should be a first consideration before applying advanced statistics, as the assumption of stationarity may pose a limitation to the applied methods from statistical point process theory. Summary characteristics that are developed for stationary patterns may lead to complex results if applied to the nonstationary patterns. We, therefore, calculated the inhomogeneous version of the $J-$function with the nonstationary intensity using kernel density estimate to assess the intensity structure of the LLS. An important issue was to determine a suitable choice of the kernel bandwidth, and to explain the effect of different bandwidths on estimating the LLS density and on the J_{inhom} functions. Apparently, a smaller-than-the-optimal bandwidth fails to reflect the inhomogeneity in the intensity surface by over-emphasising the small variations in the density surface, whereas a larger bandwidth smoothes the intensity surface, leading to the loss of important local spatial features of the density surface. From the J-functions for larger bandwidths, we observed that a larger bandwidth causes an increase of

the influence of individual LLS to a wider region and thus reduces the range of between-LLS distances.

The observed LLS pattern shows aggregated structure which may be due to interaction between the LLS, environmental heterogeneity, or both. The calculated intensity surface also reflects that the pattern exhibits large-scale variation. This should rather be attributed more to heterogeneity in environmental factors than to interaction between the locations, as in the latter case the variations in the pattern are only of a local nature. More specifically, the occurrence of a LLS in a close neighbourhood of the already existing LLS is possible and convenient due to easy access to the timber in very close proximity, hence encouraging interaction between the closeby LLS. Expansion of the logging operations over wider areas, however, is not possible without e.g. the presence of proper accessibility means such as the roads, hence necessitating the influence of the covariates on the distribution of LLS. As J_{inhom} gives an indication of inter-LLS interactions, we need to analyse the inter-LLS interactions after adjusting for the first-order effects in terms of available covariates. These covariates should be incorporated in the analysis as explanatory information to improve estimation of the density surface of the pattern. Analysis without these factors is not enough to distinguish between aggregation due to interactions or due to environmental heterogeneity. Hence, the distinction between clustering due to inter-LLS interactions and heterogeneity due to variations in environmental and ecological conditions can only be made if additional information about the environmental and ecological factors is available.

In the presence of interaction between LLS we need a suitable Markov point process to model the interactions structure. For the present study, the Area interaction model proved a suitable Markov model to detect the trends in intensity of LLS and in determination of how the trends are influenced by covariates such as the proximity to roads and clear-cut deforested patches. The model proved to be capable of incorporating not only the covariates dependence but also the interaction structure of LLS. The model provided an evidence that within the study area, the roads are a major factor determining the LLS distribution patterns, or vice versa i.e. determination of LLS in the study area may also necessitate building of roads infrastructure. Hence determining the LLS positions and building of roads are complementary processes for carryung out selective logging operations in the area.

An ultimate objective of a spatial statistical analysis of LLS is to predict the future occurrence of the spatial locations of the LLS so that the forest areas with the highest probability of degradation in the near future should receive priority attention for preventive actions and formulating effective strategies for conservation and exploitation of the forest resources. An accurate prediction of the spatial LLS pattern requires an

improvement in our knowledge of the geographic and socio-economic factors responsible for their occurrence. The land use patterns such as LLS are a result of multiple processes that act over different scales. Therefore, at different scales of analysis different driving forces have a dominant influence on determining the land use patterns. At the scale of a single Landsat scene, not much geographical variability or variations in the socio-economic data can be expected. One scene may point out the local geographic factors influencing the LLS distribution like accessibility provided by roads, as in the current study. Socio-economic factors however are usually more regional than local, whereas selective logging is not merely a local phenomena and may involve various regional factors including population dynamics, distance from important national markets, and national and governmental policies towards agriculture and forest resource exploitation. For a comprehensice analysis and accurate prediction of the future patterns of LLS distribution, an analysis of LLS at a larger scale is required. Our presented methodolgy, then, presents a quantitative method which has a great deal of flexibility to analyse the LLS patterns at multiple scales and is indispensible for analysing and interpreting LLS patterns in terms of a large number of ecological, geographic and socio-economic variables in an effective manner.

4.6 Conclusion

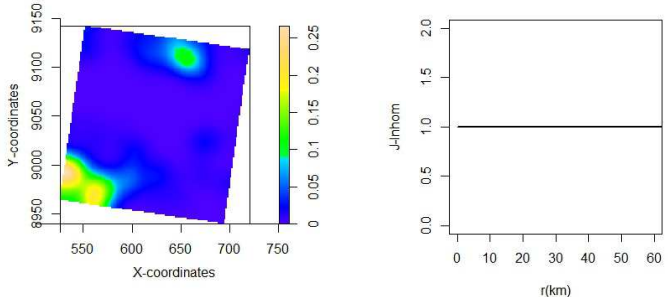
The visual and graphical methods presented in the paper provide a useful tool to get insight into the important spatial characteristics of LLS that are used as a proxy for deforestation activities in the Brazilian Amazonia. Plots of LLS reflected inhomogeneity, and the inhomogeneous J-function was used to analyse their spatial distribution. The J-function helped to infer the ranges and types of interaction using the non-parametric form of the intensity function and showed that selective logging operations are strongly aggregated.

This study leads to the following conclusions that are relevant for studies on forest degradation caused by selective logging in Amazonian forest.

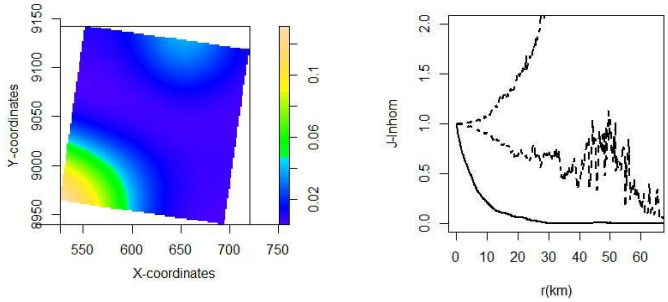
1. A spatial statistical analysis was useful to understand and interpret the LLS pattern as it occurred on a Landsat image. It showed non-stationarity and clustering in their spatial pattern that helped to determine a suitable model to represent their distribution.

2. The Area-interaction process serves as a suitable choice to model the clustered pattern of the LLS distribution in terms of the available covariates.

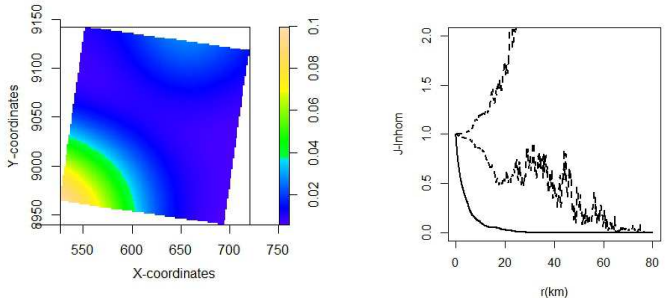
3. At the scale of one Landsat scene, roads are found to be a significant factor determining the distribution of selective logging, as well as deforestation in the study area.



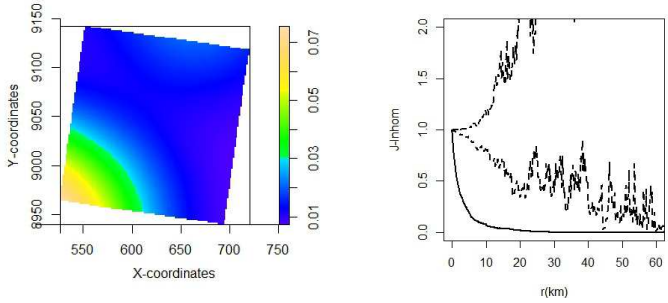
(a)Kernel bandwidth=10



(b)Kernel bandwidth=30

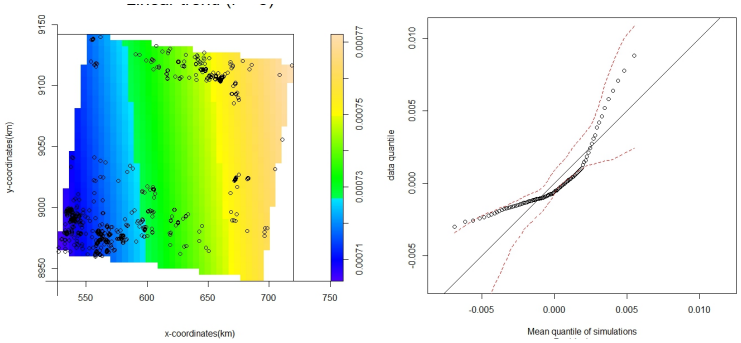


(c)Kernel bandwidth=40

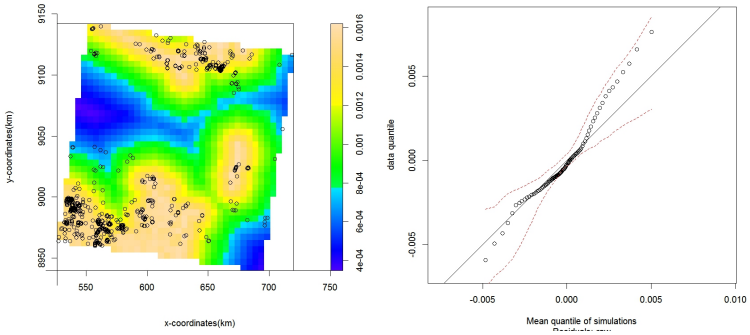


(d)Kernel bandwidth=50

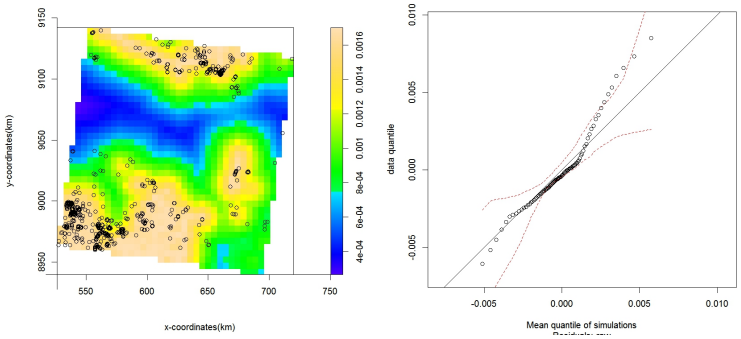
Figure 4.4: Kernel density estimates and the calculated inhomogeneous J-functions for different values of kernel bandwidth



Model 1: $\log b(x,y) = \beta_0 + \beta_1 x + \beta_2 y$ $\quad AIC = 3215$



Model 2: $\log b(x,y,DR) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 DR$
$AIC = 3210$



Model 3: $\log b(x,y,DP) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 DR + \beta_4 DC$
$AIC = 3210$

Figure 4.5: Trends fitted by different forms of the log-intensity function to the LLS pattern, along with their respective AIC values. The variable DR in the log-intensity function represents the distance of each pixel in the Landsat image of the study area to the nearest road segment, whereas DC represents the distances from the nearest clear-cut deforested polygon

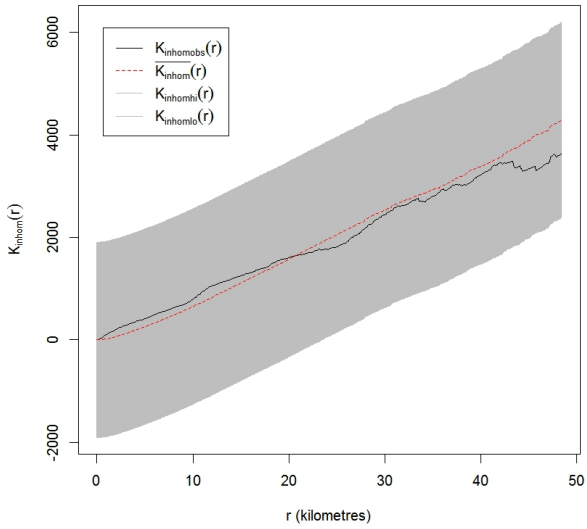


Figure 4.6: Simulated envelopes for Model 2 which estimates the intensity as a function of locations of LLS and the distance of LLS from the nearest road segment

5 Spatial pattern development of selective logging over several years

Abstract

Selective logging gives currently a major contribution to ongoing deforestation in the Brazilian Amazonia. On satellite images, log-landing sites (LLS) are well visible, and they serve as a proxy to selective logging activities. In this study we analysed the spatial patterns of the LLS collected during the years 2000-2009 in a part of the Brazilian Amazonia, using spatial statistical methods. The purpose was to reveal important spatial and temporal characteristics of selective logging. After the spatial analysis, the patterns formed by the LLS were modelled using the higher-order Gibbs interaction models due to their suitability to model clustered patterns. The Area-interaction model and Geyer's saturation model proved effective in modelling the clustered patterns in the absence of information about covariates. Results of both models conform closely to each other. We conclude that spatial statistical methods are powerful tools for analysing and interpreting the spatial patterns formed by selective logging.

5.1 Introduction

In the Brazilian Amazonian forests, selective logging for timber is a major source of forest degradation (Uhl et al., 1991). Adverse effects of selective logging on the Amazonian forests include damages to forest phenology (Koltunov et al., 2009), increasing vulnerability of a forest to fire (Gerwing, 2002; Fearnside, 2005), forest fragmentation (Broadbent

[1] This chapter is based on the following paper
Anwar, S., Stein, A., 2014. Spatial-temporal modelling of selective logging: Pattern development over several years. *Submitted to Spatial Statistics*

et al., 2008) and wide spread of deforestation (Uhl et al., 1991; Fearnside, 2005). An important challenge faced by forest researchers is to detect and analyse the locations and patterns of forest perforations caused by selective logging for timber. Selective logging operations are almost impossible to monitor as much of the Amazonian region is inaccessible. Remote sensing, due to its synoptic view and fast coverage, may serve as a viable source for the purpose. Log-landing sites (LLS), the locations where the collected timber is stored, can be detected from the remote sensing images and may serve as a proxy for the selective logging activities in a surrounding area (Anwar and Stein, 2012).

There has been intensive work on analysing the deforestation patterns and processes in the Brazilian Amazonia. Efforts have been made to model the deforestation processes using different spatial analytical tools (Apan and A., 1998; Alvis et al., 1999). Spatially explicit analysis and modelling of the locations of forest degradation caused by selective logging is, however, lacking. As logging operations vary in time, location and intensity (Matricardi et al., 2005), a spatial-temporal statistical analysis of the detected LLS may reveal important spatial and temporal characteristics of selective logging distribution. Due to small spatial extent of the LLS, it is natural to represent them as points in the maps derived from remote sensing images. Spatial point pattern statistics, then, may serve as appropriate tools for analysing and modelling the process that determines the LLS positions.

Different types of spatio-temporal data require different modelling approaches. The LLS maps generated from the remote sensing images are an example of the discrete-time spatio-temporal point process data. Discrete-time spatio-temporal point process data can arise in two ways; either the underlying process genuinely operates in discrete-time, or an underlying continuous-time process is observed at a discrete sequence of time-points (Diggle, 2003). LLS data may serve as an example of the latter as they are generated from the Landsat images acquired annually with almost the same time interval.

New logging activities grow from previously logged areas and thus extend more deeply into the interior core of remaining intact forest areas (Broadbent et al., 2008). LLS may thus exhibit interaction. Considering the interaction is important to understand the selective logging dynamics, whereas ignoring it can result in misleading conclusions about the spatial LLS distribution (Contreras-Hermosilla, 2000). If the LLS data are found to be of an inhomogeneous structure then different ecological and geographic factors may be responsible for the LLS distribution pattern. Modelling of such a nonstationary process is done using Monte Carlo Markov Chain (MCMC) methods. In the presence of interaction among LLS, such methods may become computationally extensive as they need to incorporate interaction terms as well as the spatially varying LLS

intensity (Baddeley et al., 2000). When the LLS patterns exhibit interactions, pairwise interaction point processes are suitable choice. Pairwise interaction point processes are perhaps the most widely used sub-class of Markov point processes also known as Gibbs point processes. In a pairwise point process the configuration interacts only via pairs of points. Pairwise interaction models are good models for repulsive (regular) patterns but they do not sufficiently describe clustered patterns (Law et al., 2009). The lack of pairwise Markov models for clustered patterns led to the development of higher-order interaction processes such as Geyer's saturation and the Area-interaction processes. These processes have infinite-order interactions and thus are well suited for modelling a clustered pattern.

The present study focuses on the assessment of spatial patterns formed by the LLS. Whereas the purely spatial aspects of selective logging have been dealt with in Anwar and Stein (2014b), the objective in this study is to discover important spatial-temporal characteristics of selective logging in the south-western part of the Brazilian Amazonia and then modelling the LLS patterns using a Gibbs model suitable for clustered patterns. By using patterns of several years, it is intended to better understand and quantify the spatial characteristics of the LLS distribution in the area and to reveal its temporal dynamics.

5.2 Study area and data description

The study area is described in detail in Chapter 1. Landsat 5-TM images of the study area for the years 2000, 2001, 2003, 2006, 2007, 2008 and 2009 with row 231 and path 66, were downloaded from the United States Geological Survey (USGS) website (http://earthexplorer.usgs.gov/). The images on this website are processed to Standard Terrain Correction (Level 1T), which means that the images are radiometrically and geometrically corrected. The LLS were detected from the images by applying the methodology described in Anwar and Stein (2012) which is based on the linear spectral unmixing of the images. Three fraction images were calculated representing green vegetation, shade and soil, repsectively. The soil fraction images were used to detect LLS within the forested areas.

5.3 Methodology

At the scale of the area covered by a Landsat image, the LLS can be considered as a spatial point pattern. The observed LLS data can then be described as a set of points (positions of LLS) $x = (x_1, ..., x_n)$ on a

bounded region A in the plane, assumed to be a realisation of a spatial point process with as yet unknown properties.

5.3.1 Gibbs point process

Modelling LLS data tending to show interaction, requires the use of models that allow inclusion of the interactions. Gibbs LLS processes are a natural choice in such case. The density of a Gibbs LLS process (w.r.t. the unit rate Poisson process on A) is defined as

$$f(y) = a \prod \phi(z)$$

where a is a scaling factor, guaranteeing that f is a probability density, $\phi(z)$ denotes the interaction function. The interaction function describes the type and degree of interactions between LLS, and equals 1 in the absence of interaction.

5.3.1.1 Pairwise-interaction point processes

Pairwise interaction processes are the simplest form of Gibbs processes. For the pairwise interaction processes, the conditional probability of finding a log-landing at location u given the complete realization of L is given by the conditional intensity function of the form:

$$\lambda(u, L) = b(u) \prod_{i=1}^{n(L)} c(u, x_i)$$

where $b(u)$ is the first-order term and $c(u, v)$ is the second-order or pairwise interaction term which measures dependence between LLS.

The pairwise interaction processes belong to the family of Markov (Gibbs) processes which are normally used to model the interaction structure of point patterns. These processes are adequate models for repulsive (regular) patterns but do not describe clustered patterns in a sufficient way (Law et al., 2009). An example is the Strauss model which was originally developed to model clustered patterns but later turned out to be suitable only for regular patterns. The density of the Strauss process is given by

$$f(u, L) = \alpha \beta^{n(L)} \gamma^{t(u_i, L)}$$

where $t(u_i, L)$ is the number of r- close pairs of distinct loglanding sites. β is the intensity of the loglanding process, whereas a value γ depicts the strength of interaction. If $\gamma = 1$ the model is a Poisson process with intensity β, whereas $\gamma < 1$ means that an entirely random

LLS is less likely to occur at location u if there are many LLS in the neighbourhood. Thus it reflects inhibition in a LLS process. For $\gamma > 1$, the density becomes non-integrable, hence the density is defined only for $0 \leq \gamma \leq 1$, and fails to model the clustering nature of a LLS process.

To avoid the limitations of pairwise interaction point processes in modelling clustered patterns, higher-order interaction models such as Geyer's saturation model and the Area-interaction model have been developed. Density of a higher-order interaction process depends on $m-$ tuples of points, where $m > 2$.

5.3.2 Geyer's saturation point process

Geyer's saturation model is a modification to the Strauss process. It includes an additional parameter putting an upper bound on the contribution to the density of any single point. It thus overcomes the normalizing problem in the clustered case. Due to that additional parameter, Geyer's saturation process is no more a pairwise interactions process, but a proccess of higher-order interactions. The density of the saturation process with interaction radius r and saturation threshold s can be written as:

$$f(u, L) = \alpha \beta^{n(L)} \gamma^{min\{s(L), t(u_i, L)\}}$$

with notations similar as in equation above for Strauss process.

Fitting Geyer's saturation model involves estimation of two types of parameters: regular parameters and irregular or nuisance parameters. Regular parameters like the intensity β and the interaction γ can be estimated directly from the fit. Irregular parameters like the saturation threshold s and the interaction distance r, however, have to be inferred by other methods e.g. by profile pseudolikelihood which may or may not work well (Baddeley, 2010b). For a nonstationary saturation process the value β is replaced by a function $\beta(x)$ of location x. If $s = 0$, the model reduces to the Poisson point process. If s is a finite positive number, then the interaction parameter γ may take any positive value (unlike the case of the Strauss process), with values $\gamma < 1$ describing an 'inhibitive' pattern, and values $\gamma > 1$ describing a 'clustered' pattern. If the saturation threshold s is very large, then $min\{s(L), t(u_i, L)\}$ will be equal to $t(u_i, L)$. In that case Geyer's model reduces to the Strauss model (Turner, 2009b). The interaction radius r of Geyer's saturation model should be a value in the interval from 0 to the maximum nearest neighbour distance (Baddeley et al., 2006).

5.3.3 Area-interaction point process

The Area-interaction model is also a Markov point process with interactions of infinite order. The conditional intensity function of the Area-interaction process is given by the equation

$$\lambda(u, L) = \beta\gamma^{-B(u,L)}$$

where $B(u, L) = A(L \cup u) - A(L)$ is the area of that part of the disc of radius r centred on u that is not covered by discs of radius r centred at the other $x_i \in L$. As each LLS represents the area logged within a circle of radius r, we may interpret $B(u, L)$ as the area which has not been logged yet and where a new LLS may occur. $\gamma > 1$ gives a higher chance for the specification of such a new LLS, whereas for $\gamma < 1$ a new LLS is less likely to emerge within the area represented by $B(u, L)$. Thus, $\gamma = 1$ corresponds to a Poisson process, $\gamma < 1$ produces a regular process, and $\gamma > 1$ a clustered process.

The Area-interaction process can also be described in terms of the spatial birth-and-death processes (Baddeley and Lieshout, 1995). According to this definition, we may consider L as a LLS-generating process in which the existing LLS have exponential lifetimes, since the spatial signature of the LLS disappears within two to three years. Thus, a new LLS is determined at location u with rate $p(x \cup u)/p(x)$ related to the area accessible to u that is not already reached and logged from an existing LLS.

5.3.4 Model diagnostics

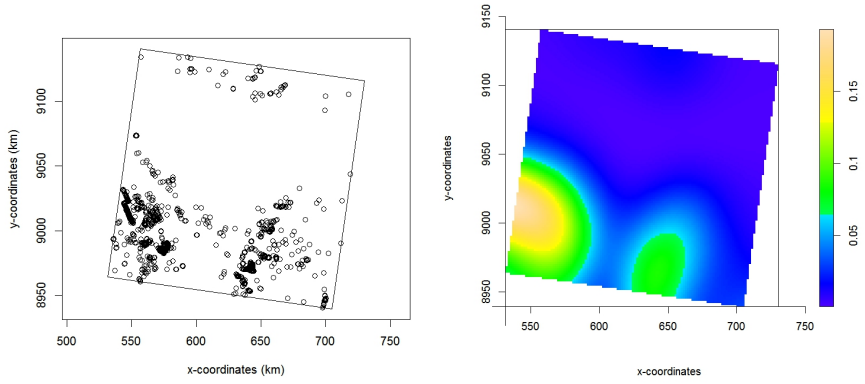
Various formal and informal methods to check the goodness-of-fit of a fitted point process model are available (Baddeley, 2008). Formal methods include χ^2 goodness-of-fit tests and Monte Carlo tests, whereas the informal methods include the residual analysis of a fitted point process model (Diggle, 2003). For Gibbs models, χ^2 goodness-of-fit test and Monte Carlo tests are theoretically not supported. Instead, the common K- and G-functions are used to simulate the critical envelopes for the fitted models which are then used as tools for checking the adequacy of the fitted Gibbs models (Anwar et al., 2011). Appropriateness of the inter-LLS interaction function in fitted Gibbs model is determined using an 'informal' validation tool known as Q-Q plot of the residuals (Baddeley, 2010a). It compares empirical quantiles of the smoothed residual field to the corresponding expected empirical quantiles under the fitted model (Baddeley et al., 2005).

5.4 Results

We first analyze the spatial LLS data from 2006 and 2007, this is followed by the data from all the years when data were collected.

5.4.1 Intensity and J-function

Figure 5.1(a) shows the locations of 990 LLS detected in year 2006. Kernel density estimate of the LLS intensity with kernel size equal to 20 km is presented in Fig. 5.1(b).

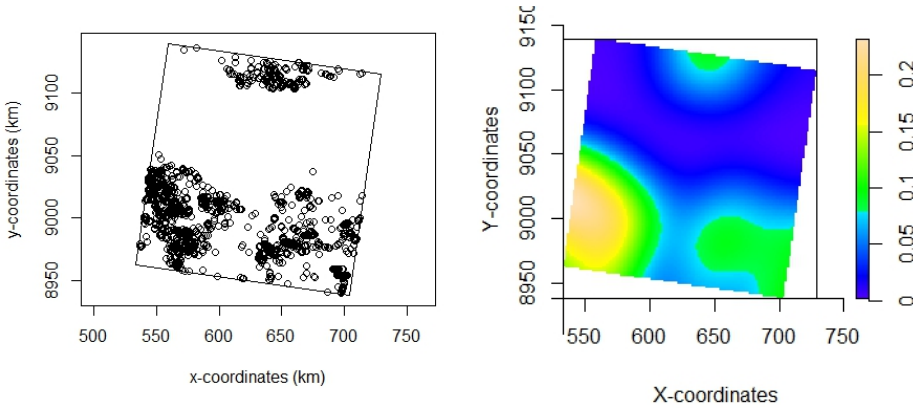


(a) (b)

Figure 5.1: (a) Loglanding sites detected in 2006, (b) kernel density estimate of intensity with kernel size equal to 20 km

Figure 5.2(a) shows the locations of 1829 LLS detected in 2007. Kernel density estimate of the LLS intensity with kernel size equal to 20 m is presented in Fig. 5.2(b).

To obtain a model that fits the LLS pattern well, we investigated the interaction structure of the LLS distribution. As Figure 5.1(b) and 5.2(b) depict an inhomogeneous structure of the LLS distribution over the study area, we calculated the inhomogeneous $J-$ function, a summary function suitable to assess the inhomogeneity, for the LLS patterns of both years. Figures 5.3 and 5.4 show the inhomogeneous $J-$ functions for the LLS-2006 and LLS-2007 respectively along with the upper and lower envelopes from simulations of a point pattern from a randomization of the data. The observed inhomogeneous $J-$ functions clearly lie below the simulated envelopes of the theoretical functions. This is evidence of



(a) (b)

Figure 5.2: (a) Loglanding sites detected in 2007, (b) kernel density estimate of intensity with kernel size equal to 20 km

a clustered pattern of the LLS in both years and we thus continue with Geyer's model and the Area-interaction model.

5.4.2 Geyer's saturation model for the LLS data

5.4.2.1 2006

To model the clustered pattern of LLS-2006, we first fitted Geyer's saturation model due to its ability to take into account the clustering nature of a pattern. For fitting this model we need to specify the irregular parameters, based upon the pseudo-likelihood estimation technique. The pseudo-likelihood estimates of these parameters were calculated and the profile log pseudo-likelihood was plotted in figure 5.5.

From figure 5.5 we observe that the log-pseudolikelihood attains its maximum value for $r = 5$ and $s = 14$. To test how well the model with these estimated values of parameters fits the LLS data, we calculated and plotted the inhomogeneous $K-$summary function for the LLS data along with the envelopes of the same summary function calculated for 99 simulated realisations of the fitted model in figure 5.6.

Figure 5.6 shows a substantial lack of fit particularly within distance between 10 and 30 km of a typical LLS of our dataset with the K-function clearly outside the range of values of the K-function for simulated patterns. The plot suggests that the LLS are more clustered within this range than suggested by our choice of parameters. Hence $r = 5$ and

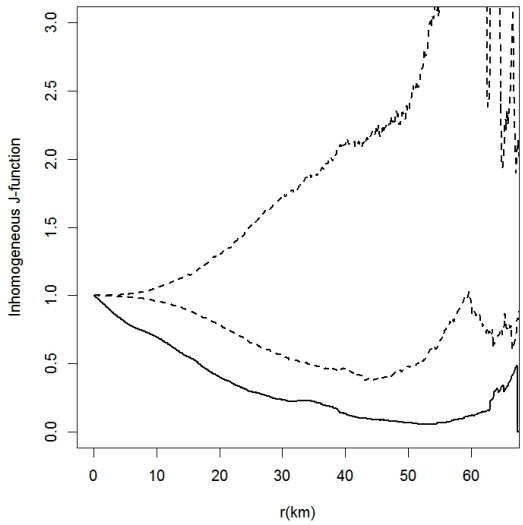


Figure 5.3: Inhomogeneous J- Function calculated using kernel density estimate with kernel size=20 for LLS-2006

s = 14 do not serve as optimal choices for the Geyer's model to the 2006 data and we conclude that log-pseudolikelihod estimation does not work well for those data. In such a case, the alternative strategy to estimate the parameter values is to consider all possible values of 'r' and 's' within a reasonable range and their combinations for fitting the model and then testing goodness-of-fit of all the models fitted thus to select the best fitting model for the LLS data. For this purpose we defined the range of possible values for the two parameters. Figure 5.3 shows that the effective range of r does not go beyond r = 30 km. As its value increases beyond its effective range, the simulated envelopes span over a wider range and the relative noise in the simulated envelopes also increases. Hence the range of r can be inferred from the inhomogeneous J-functions. Moreover, we calculated the maximum nearest neighbour distance in the LLS pattern which was found to be equal to 28 km. Therefore, we considered values of r from 5 to 25 km at steps of 5 km for fitting the Saturation model. For the saturation threshold s, low values are preferred since these make the evaluation of the log-pseudolikelihood computationally fast. We considered all values of s from 5 to 20 at steps of 5 and Geyer's model was fitted to the LLS data using all combinations of r and s within these ranges. To decide

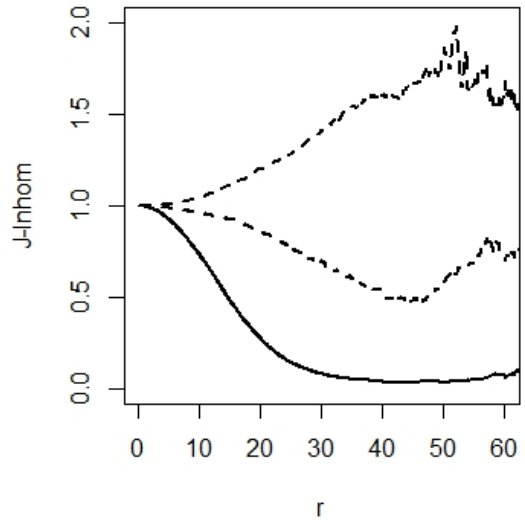


Figure 5.4: Inhomogeneous J- Function calculated using kernel density estimate with kernel size=20 for LLS-2007

upon the form of spatial trend which is most suitable to capture the remaining variability in the LLS distribution, AIC values were calculated for different forms of trend functions and compared. The intensity function which is log-linear in Cartesian coordinates was found to be the best choice.

Table 5.1 shows the estimated parameters of fitted intensities for the different models along with their AIC values, using the intensity function that is log-linear in the Cartesian coordinates.

From table 5.1 we first observe that the value of the interaction parameter γ decreases and tends to one if the value of s increases and r is kept constant, i.e. the model reduces to a Poisson process. This is because γ gives us the probability of the number of LLS neighours greater than or equal to the saturation threshold s. In other words, for large values of s, the value of γ decreases because there is a lower chance than for a LLS to have more than 's' neighbours. In that case we can not determine the true intensity of clustering in a particular data set. Therefore the value of s is kept to a minimum so as to maintain the effectiveness of Geyer's model in accounting for a clustered nature of the

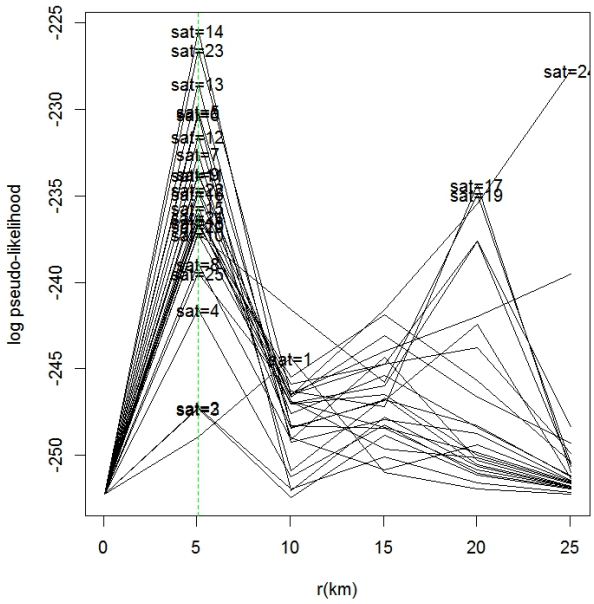


Figure 5.5: Profile-pseudolikelihood for LLS-2006

LLS data and describing the strength of its clustering in numerical terms. Secondly, we observe that as the value of r increases to the maximum nearest-neighbour distance in the LLS pattern, which in this case was equal to 28 km, the r takes large values as compared to the previous ones, indicating the presence of strong clustering. Such values of r do not appear to be credible and could be just anomalies or numerical artifacts, and therefore were disregarded.

Next, to find the optimal values of r and s, we performed a goodness-of-fit test of the above fitted models. For that purpose, the inhomogeneous K-summary functions were calculated along with the envelopes of the same function for 39 realisations of the fitted models. The inhomogeneous $K(u)$-function determines the expected total weight of all random LLS within a distance u of a typical LLS of the process. From the simulated envelopes, we found that the model with log-linear trend and the values $r = 15$ km and $s = 5$ are adequate choices to represent the LLS distribution. For this model, γ was estimated to be equal to 1.49 thus providing an evidence of moderate clustering in the LLS distribution. The K-function for the model and its critical envelopes are plotted in figure 5.7.

From figure 5.7 we observe that the inhomogeneous K-function of

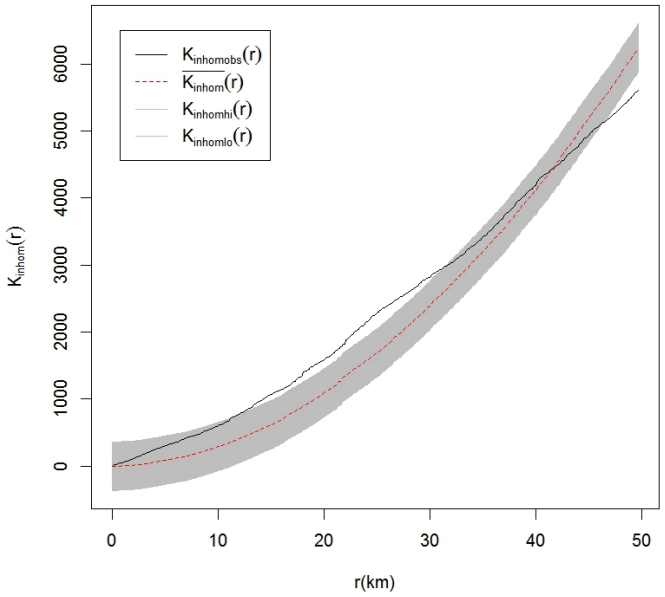


Figure 5.6: Profile-pseudolikelihood for LLS-2006

the data set calculated over a range of interaction distance lies well within the critical envelopes of the summary function under the fitted saturation model. We thus conclude that the our chosen model is a suitable fit to the LLS-2006.

5.4.2.2 2007

To model the clustered pattern of LLS-2007, we again first considered Geyer's saturation model. From figure 5.4 we observe that the effective range of r does not go beyond $r = 20$ km. Also the maximum nearest neighbour distance in the LLS pattern was equal to 16 km. Therefore, for fitting the Saturation model, we considered r values from 5 to 20 km at steps of 5 km and considered s values from 5 to 20 with steps of 5. Geyer's model was fitted to the LLS data using all different combinations of r and s within the considered ranges. AIC values of the models with different forms of trend functions were compared to determine the form of spatial trend that is most suitable to capture the remaining variability in the LLS distribution. Log-linear function of the Cartesian coordinates was found to be the best choice.

Table 5.2 shows estimated parameters of the log-linear intensity functions for different models along with their AIC values.

Interaction radius r	Saturation threshold s	AIC	Interaction parameter γ
5	5	5400	1.40
10	5	4174	1.23
15	5	2090	1.49
20	5	1064	> 10
25	5	517	> 10
5	10	5185	1.21
10	10	4110	1.12
15	10	2100	1.16
20	10	1062	1.67
25	10	516	> 10
5	15	4954	1.16
10	15	4028	1.10
15	15	2117	1.06
20	15	1044	1.70
25	15	516	> 10
5	20	4819	1.13
10	20	3897	1.10
15	20	2109	1.04
20	20	1076	1.10
25	20	514	> 10

Table 5.1: Estimated values of the interaction parameter γ with respective AIC values of Geyer's saturation model calculated for different values of interaction radis r and satuation threshold s for the LLS-2006

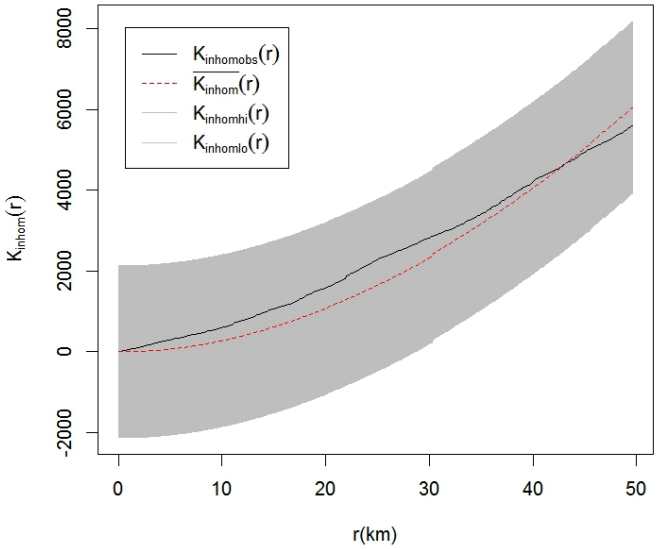


Figure 5.7: Simulation envelopes of the fitted saturation model.

Interaction radius r	Saturation threshold s	AIC	Interaction parameter γ
5	5	5400	1.48
10	5	4174	1.63
15	5	2090	1.90
20	5	1064	> 10
5	10	5185	1.24
10	10	4110	1.30
15	10	2100	1.08
20	10	1062	> 10
5	15	4954	1.15
10	15	4028	1.18
15	15	2117	1.15
20	15	1044	> 10
5	20	4819	1.13
10	20	3897	1.13
15	20	2109	1.14
20	20	1076	> 10

Table 5.2: Estimated values of the interaction parameter γ with respective AIC values of Geyer's saturation model for different values of interaction radis r and saturation threshold s for the LLS-2007

Similarly as for the 2006 data, we observe from table tab:table2007-GS that as we increase the value of s, keeping r constant, the value of interaction parameter γ decreases and tends to one i.e. the model tends to a Poisson process. Therefore we chose the minimum value of s for Geyer's model to take account of clustering in the LLS data. Secondly, as r gets closer to the maximum nearest-neighbour distance in the LLS pattern, the interaction parameter takes abnormally large values as compared to the previous ones. We ignore the models with such values of interaction parameters considering them as incredible.

Next, to find the optimal values of r and s, we performed goodness-of-fit test of all the fitted models above. For this, we calculated the inhomogeneous K- summary functions along with their envelopes for 39 simlated realisations of the fitted models. From the simulated envelopes, we found that the model with a log-linear trend and the values $r = 15$ km and $s = 5$ is a good choice to represent the LLS distribution. For this model, interaction parameter is equal to 1.90 which provides an evidence of clustering in the LLS distribution. The corresponding K-function and its critical envelopes are plotted in figure 5.7.

From Fig. 5.8 we observe that the inhomogeneous K-function of LLS-2007 data lies well within the critical envelopes of the summary function under the fitted saturation model, therefore we conclude that our chosen model is a suitable fit to the LLS-2007 data.

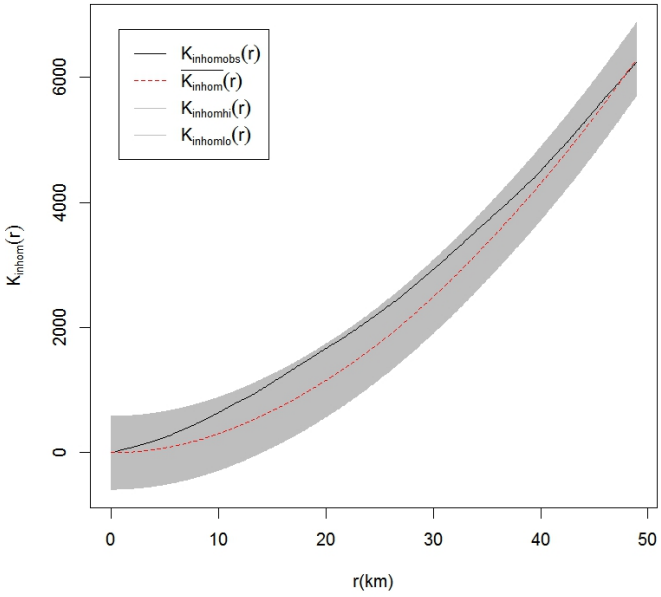


Figure 5.8: Simulation envelopes of the fitted Geyer's saturation model.

5.4.3 2000 - 2009

The above described analysis and modelling approach was adopted for the LLS data of all the years between 2000-2009 for which we had data. Table 5.3 provides a summary of the final fitted models.

Year	Number of LLS	Interaction radius r	Saturation threshold s	AIC	Interaction parameter γ
2000	650	5	5	3531	1.52
2001	917	10	5	2687	1.59
2003	752	5	5	4140	1.45
2006	990	15	5	2090	1.50
2007	1829	15	5	3509	1.90
2008	1312	20	5	1940	2.09
2009	2301	15	5	4215	1.65

Table 5.3: Estimated values of the interaction parameter γ with respective AIC values of the Geyer's saturation model calculated for different values of interaction radius r and saturation threshold s for the LLS data of the years between 2000 and 2009 for which data were available

Table 5.3 shows that the LSS intensity, as manifested by the number of LLS, is lower between 2000 and 2006 as compared to the successive years. It reflects that selective logging activities increased after 2006. In 2007 the number of LLS almost doubled as compared to the number

of LLS in 2006. The number of LLS remains high in the latter years as well. Effect of the increasing intensity can be seen in the estimates of the interaction radius for different years and the associated interaction parameters. For 2000-2003, the estimated interaction radii and the interaction parameters are smaller as compared to the ones for years 2006-2009. This means that for the years 2000-2003, log-landing activities were mainly confined to smaller areas with less inter-LLS interactions i.e. the smaller interaction radii provide evidence of smaller cluster sizes of LLS for the given years. Smaller interaction parameters with smaller interaction radii prove that the LLS are less tightly clustered over regions of smaller sizes as compared to the later years when we find that the interaction radii increase as well as the interaction parameters, indicating that the LLS are largely and more tightly clustered. For each year, the LLS detected in the previous years and repeatedly detected in the given year were taken out to restrict the analysis for a given year only to the new LLS detected in that year. Since for the years 2003 and 2006 we do not have data available for all the previous years, results for those years must be interpreted with care keeping in mind that the LLS data for these years may include some data from the previous years as well.

5.4.4 The Area-interaction model

For the Area-interaction model, pseudo-likelihood profile could not be calculated using the Spatstat software because of the memory allocation problem. Therefore, to estimate the optimal value of r of the Area-interaction process, we fitted Area-interaction model to the LLS data using r values of 1 to 10 km with a step of 1 km and tested goodness-of-fit. For each value of r we used a different forms of the intensity function and compared AIC values of the models. The intensity function that is log-linear in the Cartesian coordinates was the best choice since the higher-order polynomials did not result in significant improvement in the AIC value, and quadratric or cubic terms in the polynomials were all very close to zero. Moreover values of the interaction parameter η remained almost unaffected when including higher-order polynomial terms. Table 5.4 shows the AIC values and the estimated interaction parameter γ for all different values of r using the intensity function that is log-linear in the Cartesian coordinates.

From the table we observe that the value of γ decreases and tends to one as r increases. This means that the LLS-generating process tends to be a Poisson process, lacking interaction between the LLS, as the influence zone of each LLS increases. To determine which combination of r and γ best represents the LLS distribution, we checked the goodness-of-fit of the models for all different values of γ using the inhomogeneous

Interaction radius r	AIC	Interaction parameter γ
1	5040	3.87
2	4602	1.50
3	4652	1.22
4	4726	1.14
5	4771	1.09
6	4597	1.07
7	4411	1.06
8	4197	1.05
9	4034	1.04
10	3856	1.04

Table 5.4: Estimated values of the interaction parameter γ with respective AIC values of the Area-interaction model calculated for different values of interaction radis r for the loglanding sites data of year 2006

K-function, including the envelopes of 19 realisations of the fitted models. From the plots of the functions we observed that all calculated K-functions lied well within the critical envelopes. In such a situation it was difficult to determine the optimal value of r or the best fitting model. Therefore, we investigated the residuals from the fitted models using the QQ-plot, being a commonly used plot for Gibbs models. We found that the QQ-plot (Fig 5.9) for $r = 2$ km was showing the best fit, although still not the perfect one, for the LLS-2006.

Next, we fitted Area-interaction model to the LLS-2007 data for values of r from 1 to 10 km with a step of 1 km. For each value of r we used different forms of spatial trend. We compared AIC values of the models to determine the form of spatial trend most suitable to represent the variablity in LLS data. The intensity function which is log-linear function of Cartesian coordinates was found to be the best choice. Table 5.5 shows the AIC values and the estimated interaction parameter γ for all different values of r using the log-linear intensity function.

Interaction radius r	AIC	Interaction parameter γ
1	7947	4.02
2	7124	1.53
3	7335	1.25
4	7243	1.17
5	7123	1.14
6	4597	1.12
7	4411	1.10

Table 5.5: Estimated values of the interaction parameter γ for different values of interaction radius r of the Area-interaction model fitted to LLS-2007

To select the model that best represents the LLS distribution, we

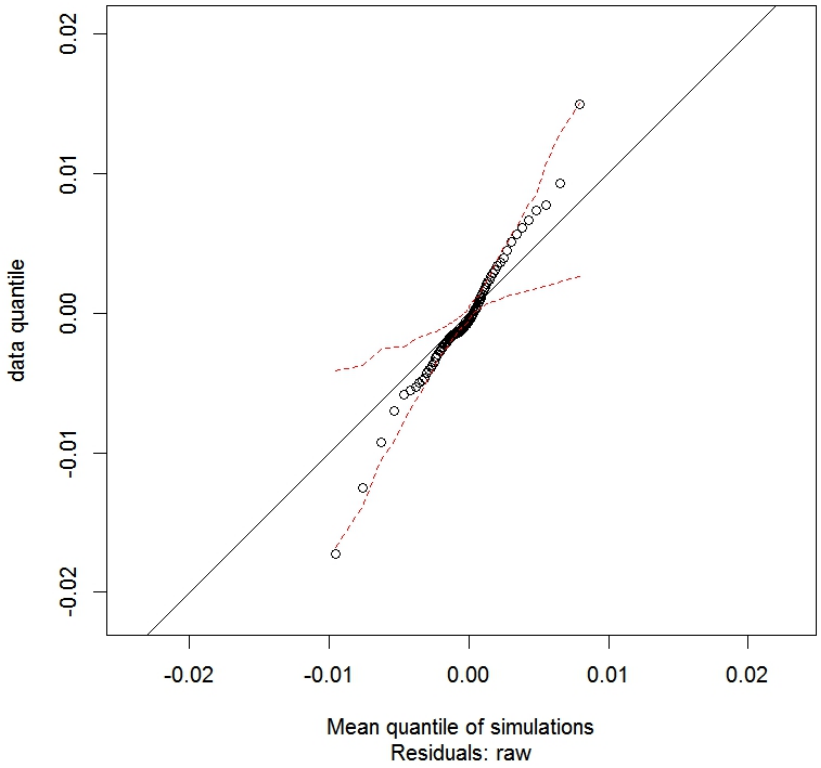


Figure 5.9: QQ-plot for the Area interaction model for LLS-2006, using r=2

checked goodness-of-fit of the models. The inhomogeneous K-functions were calculated along with the envelopes of the same function for 19 realisations of the fitted models. For this year also, the plots of the functions showed that all the calculated K-functions for the data lied well within the critical envelopes simulated from the fitted models. Therefore the simulation envelopes of the fitted models failed to provide any clues about the best fitting model. To further assess suitability of the models we investigated the residuals from all the fitted models. For this, QQ-plots of the residuals were used to determine the most optimal value of r and the corresponding best fitting model. We found that the QQ-plot (Fig 5.10) for $r = 2$ provided the best fit, although still with hint of some misspecification, for the LLS-2007.

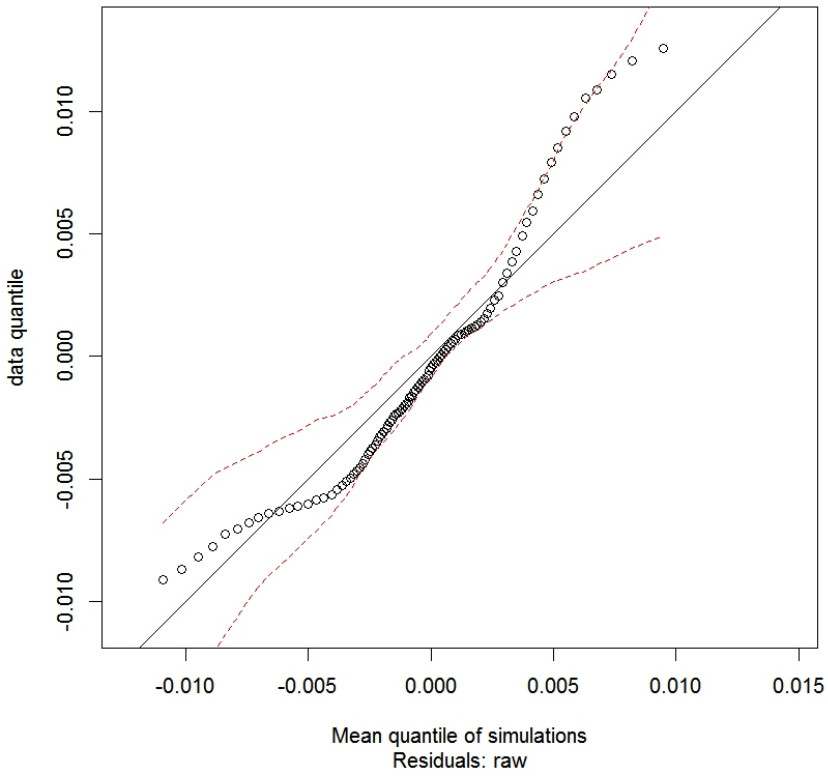


Figure 5.10: QQ-plot for the Area interaction model for LLS-2007 data, using r=3

The analysis and modelling technique described above was adopted for LLS data of all the years available. Table 5.6 provides a summary of the final fitted Area-interaction models for the years between 2000 and 2009.

From the table we observe that interaction parameters for all the years are significantly larger than 1, providing evidence of clustering in the LLS distribution of all the years. However we may note that for the years 2000-2003, the interaction radii are larger than the interaction radii for the years 2006-2009. Moreover the magnitude of the interaction parameters for the years 2006-2009 is much larger than those for the previous years. These observations reflect that the LLS in years 2006-2009 are more tightly clustered than previous years. Hence

Year	Number of LLS	Interaction radius r	Interaction parameter γ
2000	650	3	1.25
2001	917	3	1.17
2003	752	3	1.25
2006	990	2	1.50
2007	1829	2	1.53
2008	1312	2	1.46
2009	2301	2	1.51

Table 5.6: Estimated values of the interaction parameter γ with respective AIC values of the Area-interaction model calculated for different values of interaction radius r for the LLS data of year 2000-2009

the generation of LLS has changed, which may indicate a change of deforestation policies.

5.5 Discussion

Spatial point pattern statistics have been used to characterize spatial point patterns in forestry science. To the best of our knowledge, such models have not been applied to describe the patterns formed by spatial signatures left after selective logging of forests for timber. In this paper and in Anwar and Stein (2014b), we have explored point pattern techniques to characterise the LLS patterns.

We applied the Area-interaction model and Geyer's saturation process to investigate the spatial distribution of LLS. The Strauss process and the Area-interaction process are examples of point processes from two large classes of processes: the distance-interaction processes and the shot-noise weighted processes. We have shown the effectiveness of both types of processes in modelling clustered patterns such as the LLS distribution in our study area. In general, both approaches worked well for our data and the results of both models are close to each other. Simulation envelopes and a QQ-plot were used as diagnostics for model validation. Although the QQ-plots showed that model fitting was satisfactory, they did not show a perfect fit. This does not necessarily mean that the interaction structure was mis-specified, as the discrepancies in a QQ-plot may result from a number of uncontrolled factors such as mis-specification in the trend function or presence of outliers. It was also observed in this study that the QQ-plot is sensitive to slight mis-specifications in the range of interpoint interactions. Therefore different values of interaction distance with small intervals must be tested. This, however, may not always be feasible due to large number of possible values within a given range and the large amount of time it requires to calculate QQ-plot for each value.

Correct interpretation of QQ-plot also depends upon the correct form of spatial trend and intensity function used. In case of an inhomogeneous LLS pattern, even if the spatial trend is correctly determined, the intensity function may involve dependence on covariates. Such dependence may or may not be observable and it may thus be difficult to measure and incorporate all of the covariate effects into a fitted model. Hence the QQ-plot may show lack of fit resulting in misleading conclusions about the fitted model. Specifically, the distribution of LSS is believed to be determined by a number of local and regional socio-economic and environmental factors. At the scale of a single Landsat scene, it is not possible to measure the effect of all the factors involved and hence the intensity may not be fully modelled in terms of covariates.

In practice, Gibbs models currently available for use in data analysis are low in number and limited in scope. In real world situations, however, we come across a plethora of different examples of point patterns with different distributional characteristics. The available set of spatial statistical models is therefore insufficient to represent this diversity. It can not be expected that one or other of the developed models will appropriately apply to a given situation. Due to diversity of real-world situations, the scope of available set of models might need to be expanded. Limitations in doing so concern the formulation itself of such a model and fitting it to the data set may involve skillful programming, apart from the more technical and theoretical prerequisites such as integrability and local stability of the functions used. Moreover distinguishing between existing models is sometimes difficult, in particular as only the data and not detailed information of the process involved in the data generation is available to guide the choice of suitable model.

Despite the limitations stated above, our fitted models proved to be satisfactory to model the distributional characteristics of the LLS data, in particular from the diagnostic tools used. In a forestry context, our study can provide us with an overview of the characteristics of LLS data in spatially explicit and quantitative terms. Our study has presented several spatial point pattern methods which are, we believe, of importance to the forest statistician. The LLS distribution reflects the location and density of valuable timber species in a study area and may point to possible degradation. Forest degradation caused by selective logging, however, is a complex process affected by a plethora of socio-economic and environmental factors. With the use of covariate information, it is expected that the fitted spatial statistical models can give us improved results which can be used for predictive purposes to determine the future trends in the distributon of selective logging. Further, the change in parameter values over the years has given some idea in changes in deforestation activities. The stronger clustering points to an intensified deforestation, where new LLS are located at shorter

distances than before. Such information is useful for forestry studies and may be lucrative for timber loggers in order to carry out selective logging operations in a cost- and time- effective manner.

In the future, the results of this study could be used to investigate the forest structure left after the logging operations. This information may also be required by the forest conservation agencies to plan effective strategies for exploiting the forest resources without causing significant damage to its canopy structure and composition of its flora and fauna. The spatial point pattern methods may be included into the everyday toolbox of exploratory methods for analyzing the selective logging distributions in the Amazonian forests.

5.6 Conclusions

The spatial statistical methods presented in the paper provide a useful tool to get insight into the important spatial and temporal characteristics of LLS that are used as a proxy for deforestation activities in the Brazilian Amazonia. Plots of LLS reflected inhomogeneity, and the inhomogeneous J-function were used to analyse their spatial distribution. The J-function helped to infer the ranges and types of interaction using the non-parametric form of the intensity function and showed that selective logging operations are strongly aggregated. Geyer's saturation process and the Area-interaction process were shown to be effective in modelling a clustered pattern such as selective logging distribution in the study area.

This study leads to the following conclusions that are relevant for studies on spatial statistical analysis of forest degradation caused by selective logging in Amazonian forest.

1. A spatial statistical analysis was useful to understand and interpret the LLS pattern as it occurred on a Landsat image. It showed non-stationarity and clustering in their spatial pattern that helped to determine a suitable model to represent their distribution.
2. Geyer's saturation model and the Area-interaction process model serve as suitable choices to model the clustered pattern of the LLS distribution even in the absence of any information about covariates information.
3. Results of Geyer's saturation process model and the Area-interaction process model conform closely to each other hence both models prove to be effective in representing the clustered LLS pattern.

Synthesis

6

6.1 Introduction

This chapter reviews the main elements of the thesis. It summarizes how aims and objectives of the thesis were achieved and to what extent they were successfully accomplished. It also provides suggestions and recommendations for future extensions of this research.

Detailed conclusions for the individual studies have been presented in the individual chapters. In this chapter I summarize the general conclusions from the work and discuss the key findings of this research in relation to the research objectives.

The main objective of the research was to investigate the potential of spatial point pattern statistics as a tool for spatially explicit analysis and modelling of initial patterns of land cover changes. The objective was accomplished by investigating the contribution of anthropogenic influence and the natural factors to LULC changes separately. The research comprises two case studies:

1. For investigating the LULC change patterns due to natural factors, we chose earthquake epicenter locations as a source of LULC changes due to natural forces.
2. For investigating the patterns of LULC changes due to human influence we analyze the distribution of log landing sites (LLS) as a source of LULC change resulting in deforestation of the tropical forests.

For both studies, we started by analyzing the distribution of initial sources of LULC changes. Both analyses aimed at detecting important spatial characteristics of the distributions of initial LULC and then identifying suitable point process models for each study based on the detected spatial characteristics. Since the aftershocks of an earthquake occur instantaneously in time, we study only the spatial aspects of the earthquake aftershocks data. On the other hand, locations of the LLS form trajectories (sequences of successive events) in space and time,

and since we have temporal data available for the LLS distribution, it was possible to investigate the temporal trends in the spatial distribution of LLS in order to get further insight into the processes responsible for the LLS distribution. These tasks were the sub objectives of the research.

6.2 Research objectives

In the following paragraphs, I present a summary of the results of each of the objectives stated in section 6.1 stating their importance and implications.

6.2.1 Research objectives 1: Implementation of the marked Strauss point process model to the epicenters of earthquake aftershocks

Modelling earthquake data has since long been a focus of research by seismologists and statisticians. A spatial analysis of earthquakes seismic locations serves as a first and important step to give insight in to the underlying geological processes. The earthquakes locations are found in forms of clusters on the globe and their distribution is determined by many observed and unobserved geological and geophysical characteristics of the epicenter region. Modelling the data constituting the locations of earthquake epicenters requires some model that can best represent the clustering behavior of the earthquakes by take into account the effects of the external factors. Having a better knowledge on where the earthquakes, e.g. as major events or as aftershocks occur in relation to geological features, may thus result in identification of hazard zones. This study explored different spatial statistical techniques to analyze earthquake data in Pakistan recorded since 1973, including a major event in 2005. The Strauss point process model was investigated for its flexibility to incorporate available geological information such as the presence of faults and plate boundaries as explanatory variables, and for its appropriateness to model this marked and clustered pattern.

Specific findings related to this objective were:

1. The Strauss model proved to be flexible and rigorous in modelling the clustered pattern of the earthquakes and it could explain the variability of earthquake locations distribution.
2. With the inclusion of explanatory variables, the results of Strauss model fitting became increasingly refined; hence it has the capacity to incorporate a variety of additional variables to define a marked point pattern such as earthquake epicenter locations data.

3. The modelling approach suggested that the places close to the plate boundaries are associated with high probability of earthquakes occurrence.
4. Locations of active tectonic faults also proved to be a very significant determinant for the distribution of earthquake epicenter locations as incorporation of the information about the fault locations greatly improved the model fitting results.

6.2.2 Research objectives 2: Detection and spatial analysis of selective logging as a source of forest LULC

The Brazilian Amazonian rain forests are under imminent threat of serious degradation and ultimately deforestation. Human activities such as selective logging are an important cause. Selectively logged forests are difficult to detect from medium-resolution Landsat images, due to their relatively small sizes and subtle spatial patterns. Spectral linear unmixing provides an effective tool for the purpose. After detection of selectively logged locations, we need to analyze the locations in a spatially explicit manner to gain insight into the distributional characteristics of the selective logging in the Amazonian forests. Understanding the pattern of log-landing sites is a first prerequisite in any effort to model the distribution of log landings which in turn may help us to develop effective strategies to conserve the tropical forests and maintain its biodiversity. Spatial statistical analysis of the detected log landing sites showed a strong clustering within the study area. We conclude that the endmembers used in this study represent basic components of a degraded forest environment. As spectral unmixing of remote sensing images avoids collection of field data, it may broadly be applied towards other Amazonian regions as well. This study leads to the following conclusions that are relevant for studies on deforestation in Amazonian forest with a selective logging detection:

1. The endmembers used in this study are well suited to identify selective logging locations. They represent basic components of a degraded forest environment.
2. For the currently available remote sensing images with a resolution of 30 m, spectral unmixing is a well-suited technique to identify patterns of selective logging locations. It may either avoid the collection of field data or help to reduce sampling efforts.
3. The image rotation has been suggested to be a better approach since it offers the advantage of selective logging detection over a larger area. Positional errors up to 30 m, however, have to be compromised.

6.2.3 Research objectives 3:Use of spatial statistics to investigate early forest degradation activities detected from satellite images

Selective logging gives currently a major contribution to ongoing deforestation in the Brazilian Amazonia. The spatial distribution of log landing sites (LLS), i.e. the sites where logged trees are collected, serves as a proxy to the intensity of selective logging activities. In this study we analyzed the LLS pattern in a study area that has a rapid deforestation. Actual LLS locations were extracted from a Landsat image of 2000 that covers a large part of the study area. We first used the inhomogeneous J-function. A kernel bandwidth of 20 km best modelled the non-stationarity, showing a strongly clustered LLS distribution. Second, the Area-interaction point process model incorporating information about distance of LLS to roads and to clear-cut deforested areas was applied. The model well explained the clustered LLS pattern and showed a significant effect of distance to roads. We concluded that this spatial statistical study helped to quantify and better understand the LLS pattern. This study leads to the following conclusions that are relevant for studies on forest degradation caused by selective logging in Amazonian forest.

1. A spatial statistical analysis was useful to understand and interpret the LLS pattern as it occurred on a Landsat image. It showed non-stationarity and clustering in their spatial pattern that helped to determine a suitable model to represent their distribution.
2. The Area-interaction process serves as a suitable choice to model the clustered pattern of the LLS distribution in terms of the available covariates.
3. At the scale of one Landsat scene, roads are found to be a significant factor determining the distribution of selective logging, as well as deforestation in the study area.
4. This study also supports the argument that deforestation is an inertial process by which the areas most likely to be deforested are those located closer to the forest areas already intervened.

6.2.4 Research objectives 4

Spatial-temporal modelling of selective logging 2: Pattern development over several years

Selective logging gives currently a major contribution to ongoing deforestation in the Brazilian Amazonia. On satellite images, log-landing sites (LLS) are well visible, and they serve as a proxy to selective logging activities. In this study we analyzed the spatial patterns of the LLS

collected during the years 2000-2009 in a part of the Brazilian Amazonia, using spatial statistical methods. The purpose was to reveal important spatial and temporal characteristics of selective logging. After the spatial analysis, the patterns formed by the LLS were modeled using the higher-order Gibbs interaction models due to their suitability to model clustered patterns. The Area-interaction model and Geyer's saturation model proved effective in modelling the clustered patterns in the absence of information about covariates. Results of both models conform closely to each other. We conclude that spatial statistical methods are powerful tools for analyzing and interpreting the spatial patterns formed by selective logging.

This study leads to the following conclusions that are relevant for studies on spatial statistical analysis of forest degradation caused by selective logging in Amazonian forest.

1. A spatial statistical analysis was useful to understand and interpret the LLS pattern as it occurred on a Landsat image. It showed non-stationarity and clustering in their spatial pattern that helped to determine a suitable model to represent their distribution.
2. Geyer's saturation model and the Area-interaction process model serve as suitable choices to model the clustered pattern of the LLS distribution even in the absence of any information about covariates information.
3. Results of Geyer's saturation process model and the Area-interaction process model conform closely to each other hence both models prove to be effective in representing the clustered LLS pattern.

6.3 Reflections on the study

In a spatial statistics context, the two chosen examples as sources of LULC change represent the two types of space-time point processes, namely space-time shock point process and space-time survival point process as defined by Cressie (1993). The two types can be distinguished based on the duration of LULC events over time. Space-time shock processes occur instantaneously over both space and time, whereas space-time survival processes are born at some random location and time, and then live for a random length of time. On the one hand, occurrence of earthquake aftershocks serves as a realization of space-time shock point process in which temporal distribution of the aftershock does not offer much information about the process, therefore we mainly focus on the spatial aspects of the study. On the other hand, space-time survival processes can be used to model the locations of LLS as source of LULC changes due to selective logging. Due to the possibility of data

accumulation over regular time periods, we could extend the spatial point pattern analysis of the study to also investigate temporal patterns of the LLS distribution.

6.3.1 Earthquakes as natural source of LULC change

A major goal in statistical seismology is to develop models that offer the ability to accurately anticipate future earthquakes to save mankind from the vast destructions caused by the earthquakes. Several aspects, however, need careful consideration while trying to develop such models.

First, the earthquake aftershocks data are difficult to fully explain, since the variability in the pattern is caused by many observed and unobserved geological and geophysical factors. The phenomena occurring deep down the earth's surface may not manifest itself in the form of clear indications above the earth's surface and this makes it difficult to search for possible indications or covariates associated with the spread of earthquake aftershocks and their magnitudes. For a realistic model fitting, information about the earthquake-generating mechanism is required to fully explain the process. Moreover, the process of earthquake formation is complex. Large phenomena such as the earthquakes relate to subsurface physical processes of the dynamic earth which do not evolve in countable times. Generation of the whole mechanism is spanned over a long period and therefore it could take long times, decades or even centuries, to accumulate the information on the earth's geophysical characteristics related to the earthquakes occurrences, to fully substantiate a statistical model for predicting large earthquakes.

Apart from the non-availability of reliable precursors related to the earthquake occurrences, the earthquake epicenter locations suffer from uncertainties caused by measurement errors and by the unfavorable geometry of stations that record the earthquakes. Uncertainties in earthquake locations and their magnitudes can also be introduced due to errors in station parameters (location, timing), misidentification of seismic phases, and due to errors in the velocity model used to compute the earthquake locations (Husen and Hardebeck, 2010). These errors lead to a systematic bias and hence they affect the accuracy of an earthquake data. Moreover, they are difficult to detect and do not follow standard statistical distributions, which makes it difficult to properly account for them in determining locations and magnitudes of earthquakes very accurately. Since the spatio-temporal properties of earthquakes cannot be studied directly; one has to rely on the earthquake catalogues to infer about the spatio-temporal characteristics of the earthquakes distribution. Hence accuracy of the detected spatial-temporal characteristics of the earthquakes distribution depends on the accuracy of the catalogue

used. Therefore, when interpreting results of a statistical study based on an earthquake catalog, it is vital to consider the limitations of the catalog and how these limitations may affect the study's results.

There are several other factors also that can greatly influence the output of a statistical model of earthquake. One of them is differentiating between the main shocks and aftershocks. By definition, aftershocks are smaller in magnitude than the main shock and occur within 1-2 rupture lengths distance from the main shock (ACE, 2010). The proximity and magnitude of the subsequent events is, therefore, considered very important to define the type of earthquake. Large earthquakes are able to trigger additional earthquakes, ranging from aftershocks in the immediate vicinity to remote earthquakes hundreds of miles away. Therefore occurrence of the next earthquake depends on the characteristics of the preceding events and the dynamics of the fault system on which they occur. Another factor is the time and span of the events. A main event and an aftershock could theoretically occur more than a year apart. However, the larger the main shock, the longer they will continue to produce aftershocks. Aftershocks can, thus, continue for weeks, months or even longer. We can, therefore, not yet strictly distinguish between main shock and aftershock based on the occurrence time.

6.3.2 LLS as human-induced source of LULC change

6.3.2.1 Detection of LLS

Selective logging of forests in the Brazilian Amazonia is one of the most striking examples of human-induced source of LULC change. Detection of selectively logged forests poses a serious challenge to the forest conservation efforts due to subtle spatial patterns of initial LULC changes caused by selective logging. Conventional digital image processing techniques such as visual interpretation, minimum distance and maximum likelihood classifiers have been evaluated for mapping selective logging in the Amazon. In our study we used spectral linear unmixing of remote sensing multispectral images to overcome some of the problems of visual interpretation and conventional image processing techniques. The soil fraction derived from linear spectral unmixing enhanced the detection of the log landing sites which were recognized as the spatial signature of mechanized logging in tropical forests.

Although the technique is generally applicable to the other regions in the Brazilian Amazonia, some precautions must be taken when considering other Amazonian forests for its application:

In our study on detecting the LLS, we applied some threshold values which should be specifically determined for the area under study. The first threshold was applied to identify forested pixels with high soil

fraction (i.e., Soil>20%) associated with the LLS. Next, an area threshold was applied to map the LLS (i.e. 1 to 4 contiguous pixels). These threshold values may vary depending on the forest type such as dense and open forests, and density of valuable timber species in other forests of the Amazonia. Therefore caution should be taken in determining the thresholds when the technique is applied to the other Amazonian forests.

There are a range of canopy densities and ecosystem types across the tropics. This natural variation in forest cover can be due to underlying biophysical elements (e.g. semi-arid, semi-deciduous, shrub lands and soil conditions). To many satellite sensors, some ecosystems characterized with more open canopy structure may appear similar to degraded areas of neighboring forest. Human intervention in such canopies is very difficult to distinguish. This type of confusion could be alleviated to a degree by having an accurate and detailed vegetation map of the various natural canopy types, if available, which can be used to aid contextual information and expert knowledge.

Access to cloud-free images over the Amazonia is critical for monitoring forest LULC changes. Even light clouds or haze over tropical forests can be a problem because it is often confused with degraded forests during satellite image classification. Coarse resolution imagery (e.g. MODIS) has sufficient temporal frequency for a time series (ideally multiple images per year to distinguish degradation from effects of seasonality). However, at coarse resolution much forest degradation, which is often small scale, can be missed.

In this study, we separated LLS from natural forest disturbances using contextual information. The contextual information relies heavily on the expert knowledge which is always prone to errors of omission and commission. Moreover a forest mask was applied to avoid confusing LLS with other small cleared areas found in deforested areas. Accuracy of the final map of LLS therefore depends on the reliability of expert knowledge and accuracy of the classification techniques used.

We used locations of LLS sites as proxy for selective logging in the study area. Sometimes the LLS are built very close to the roads and hence separating the LLS signals from the roads is not possible. Moreover, in the non-mechanized selective logging, loggers do not build log landings and the timber is directly extracted out of the forests. In such cases, our technique to detect and analyze the selectively logged forests may not work and hence other means to detect and analyze the selectively logged forests must be looked for and employed.

6.3.2.2 Spatial point pattern analysis and modelling

Most LULC change models attempt to untangle the driving forces behind anthropogenic sources of LULC change, including socioeconomic and biophysical driving forces. LULC change patterns nearly always exhibit spatial autocorrelation which is due to neighborhood effects in LULC change patterns. The clustered distribution of most landscape features provides evidence of neighborhood effects which may be due to gradients in environmental conditions that are important determinants of the LULC change pattern or due to the spatial interactions between LULC change features itself. In the LLS pattern context, the spatial interaction between the LLS can be explained by the fact that new LLS are determined next to the already placed LLS, as an extension to the logging operations further into the interiors of the pristine forests.

Cellular automata and regression models have been two most widely used models in forest LULC change studies. The theoretical basis of addressing and quantifying the autocorrelation in both methods is, however, considered poor (Verburg et al., 2004). It has, therefore, always been recommended by the forest LULC change modelers to develop more sophisticated ways to account for the neighborhood effects in LULC change models. The neighborhood effects must be regarded as a consequence of interactions both between neighboring land uses and between the driving forces. Ignoring any of these two sources of interactions may lead to biased results in LULC change models. Particularly if interaction between the driving forces is not properly accounted for, the model may end up lacking causality. The point pattern statistics offer the ability to account for the neighborhood effects in the LLS patterns by investigating the interaction structure between the LLS as well as by quantifying the interaction between environmental factors.

Distance-based summary functions in the point pattern statistics e.g. the inhomogeneous J-function used in our study, are able to detect the form and range of spatial interactions. The distance functions thus can be helpful to answer different questions about scales of overall pattern and about segregation (regularity) or aggregation (clustering) in the LLS patterns.

The Area-interaction model was found to be a suitable model for explaining the distribution of LLS in terms of available covariates. The model proved successful not only in assessing the contribution of the covariates information, it was also helpful in detecting and quantifying the interaction strength between the covariates given by the locations of roads and locations of deforested areas in the study area. It showed that both covariates can serve as proxy to each other due to perfect multi-collinearity existing between them.

Results from our study have practical implications for managing and

monitoring tropical forest landscapes. Our study found that within the study area, roads play a significant role in determining the selective logging. Loggers build roads to get in the interiors of forests for carrying out mechanized logging operations. The roads built for the purpose of selective logging encourage cattle ranchers and agriculturalists to invade through the forests and, in this way, the roads contribute towards deforestation of the area. More attention should, therefore, focus on closing logging roads after harvest operations (e.g. by destroying key bridges or otherwise rendering the road impassable) to inhibit post-logging invasions of forests.

6.3.2.3 Spatio-temporal modelling

Although a study of purely spatial patterns of a forest structure reveals important spatial characteristics, in real life, forests develop dynamically through time and hence study of the temporal aspects is vital to understand the forest dynamics (Comas and Mateu, 2007). Analysis of such dynamic processes requires understanding of the physical processes underlying the initiation and subsequent evolution of changes in a forest LULC.

Maps of the study areas obtained through multi-temporal sequence of remote sensing have made it possible to analyze the spatial-temporal dynamics by describing how the patterns evolve over time (Mertens and Lambin, 1997). Tracking individual trajectories i.e. sequences of successive changes in LULC, is possible if patterns found in one map, obtained through remote sensing techniques, are linked to the subsequent maps. In this way the temporal series analysis can be helpful to project the likely temporal evolution of a landscape in future. A static analysis of the spatial patterns of LULC change without taking the trajectories of change in to account may lead to erroneous interpretations of the current and future LULC change dynamics. Since LULC changes may be dependent on initial conditions and different combinations of socio-economic data at different regions may result in very different spatial patterns of LULC change, for understanding current patterns of forest, LULC must be analyzed with a combination of policy, accessibility, and biophysical and socioeconomic factors while accounting for the historical pathways of change (reference). Availability of data with the necessary temporal and spatial resolution may be the most important constraint for such research.

Since we aggregate LLS points per year, linking each year's spatial patterns is a better approach for investigating the temporal aspects of the study. Spatial-temporal studies on finding the temporal clustering in time can be done if we get time information attached to the individual LLS points which is not possible due to the temporal resolution of the

images. We can only accumulate the points per year due to availability of the images. Temporal resolution of the images is limited by the fact that visibility of the region by satellites is hindered by the cloud cover throughout most time of the year. It is only during the dry season July-August that the region is observable through satellites. Moreover, the logging operations are also carried out mostly during dry seasons because the terrain conditions in the area become less favorable for extracting and transporting timber during rainy season. Therefore rather than modelling the LLS point patterns with a spatio-temporal point process, we aimed at estimating the parameters of purely spatial point processes of LLS patterns obtained annually and exploited the information contained in the temporal replicates.

6.4 Outlook for Further Research

The process of earthquake formation is complex and depends on the subsurface physical processes of the dynamic earth which are not precisely known to humans. In future, as the geophysical earthquake generating mechanism becomes better understood, it would become inevitable to develop new models for earthquake occurrences. Future research efforts should also be concentrated on developing models which attempt to reflect the physical mechanism of seismic events even if sufficient data for developing the parameters of these models are not available. The uncertainty in these models should be addressed to reflect the systematic error due to the insufficient or inaccurate data. The uncertainty can be reduced over time as more data on earthquake parameters become available. Moreover there is growing evidence that earthquakes occur in clusters in both time and space. Scientists need to build models to explain this clustering behavior. The models can serve to predict which regions are prone to aftershocks and other earthquakes, after a major earthquake occurs. The model can then be used for preparing seismic hazard maps based on the probabilities obtained. Anticipating major earthquakes may be helpful, not only for short-term response such as preparation of emergency personnel and disaster relief, but also for longer-term preparation in the form of building codes, urban planning and earthquake insurance.

The LULC patterns such as LLS are a result of multiple processes that act over different scales. Therefore, at different scales of analysis different driving forces have a dominant influence on determining the LULC patterns. Our study analyzed the LULC pattern given by LLS at the scale of a single Landsat scene. At this scale, not much geographical variability or variations in the socio-economic data can be expected. One scene may point out the local geographic factors influencing the LLS

distribution like accessibility provided by roads, as in the current study. Socio-economic factors however are usually more regional than local and selective logging is also not merely a local phenomenon and may involve various regional factors including population dynamics, distance from important national markets, and national and governmental policies towards agriculture and forest resource exploitation. For a comprehensive analysis and accurate prediction of the future patterns of LLS distribution, an analysis of LLS at a larger scale is required. Our presented methodology can be extended further to analyze the LLS patterns at broader scales in terms of a large number of ecological, geographic and socio-economic variables in an effective manner. In a model of multivariate spatial point patterns, the number of species and interaction parameters can become very large leading to highly complex models. This necessitates a suitable modelling approach which facilitates model selection.

In LULC studies over broader areas, we may need to investigate the interaction structure of LLS at multiple scales. The existence of different causal processes at different scales means that spatial interactions should be studied at multiple scales. Area-interaction model, however, does not allow for models whose behavior changes at different resolutions, for example repulsion at small distances and attraction at large distances. The capabilities of the model need to be further expanded to account for the multiple scale issues within one LULC pattern.

Owing to several sources of possible inaccuracies/uncertainties in detection of LLS, there is still need to develop innovative methods and analytical techniques coupling satellite imagery with ground-based observations to detect this obscured form of forest degradation more accurately and precisely. This can be done through developing better pattern-recognition approaches. Moreover, since the spatial extent and distributional patterns of selective logging may vary across Brazilian Amazonia, there is need to understand the spatial and temporal heterogeneity of forest damages caused by selective logging in order to diminish the effects of selective logging on ecological and biogeochemical processes in a forest environment.

The research should be extended to construct a single model that fits the temporal replicates and that accounts for differences among the replicates, i.e. the yearly variations. It may be feasible to combine the information from all the replicates enabling the analysis of data sets in small study areas where only a few points can be observed at any given time. However for large data sets it may be more challenging. The approach would allow the simultaneous analysis of replicated patterns in one single model by considering random effects that account for between-replicate differences.

6.5 Data quality and uncertainties

In our study we have not gone in much detail about the data quality aspects and uncertainties associated with the data and hence the model outputs. Empirical studies on LULC, however, do face the obstacle of inadequate data.

In the study of LULC changes due to natural forces, different sources of inaccuracies in recording or measuring the earthquake epicenter locations and magnitudes have been discussed in detail in section 6.3.1. Care should be taken when using earthquake location catalogs in statistical seismology studies, as the catalogs are not exact representations of the true earthquake locations. Instead, earthquake locations are affected by random and systematic errors (reference). Completeness and accuracies of the earthquake catalogues must be assessed before studies aimed at more practical uses of the catalogues and for interpreting the results of models for more practical purposes such as delineating earthquake prone areas for better planning of the infrastructure.

Our study on LULC change due to human influence relies heavily on the remotely sensed images and techniques to detect the locations of LLS. As discussed in section 6.3.2, due to inherent variability in canopy structure in different ecosystems within the Amazonia, ecosystems characterized with more open canopy structure may appear similar to degraded areas and hence may cause inaccuracies in the detection of LLS using satellite images. Moreover, due to persistent cloud cover over the region throughout most of the time of the year, cloud cover and its shadow over the selectively logged forests may also hinder the detection of LLS in the region thus causing inaccuracies in the estimates of total extent of selectively logged forests. These sources errors of omission and commission need to be addressed for more practical purposes of the studies on selective logging.

Apart from the errors of omission and commission, there may be uncertainties associated with the area of each LLS detected, as they were inferred using the pixel information which depends highly on the spatial resolution of the sensor used. The greater the surface detail a sensor can observe, the more precise boundaries and areas of the LULC can be determined from it. However the tradeoff between detail and coverage makes high-resolution sensors less suitable for routine deforestation mapping on a global scale. Hence remote sensing methods have to rely on medium resolution remote sensing images which suffer the problem of identifying the LLS at smaller scale with uncertainties involved in the estimation of the area and extent of the LLS.

Remote sensing measures the reflective response of the earth's surface, and so it can be used to directly observe the land cover for a given

pixel. Unlike land cover, which can be directly observed and monitored from remote-sensing data, land use typically must be inferred through a combination of remote-sensing observation, regional and local knowledge (including field observation), and other ancillary information that links a given land cover in a region with a given land use. In the absence of field data, we had to rely on expert knowledge to link the forest land cover patterns in our study with the land use activity. Expert knowledge however is also prone to errors and thus may misclassify some land cover.

Bibliography

ACE, 2010. Cat 360. catastrophe risk from every perspective: Making sense of earthquake clusters. Web site, accessed on 18 July 2014.

Adam, J., Sabol, D., Kapos, V., Filho, R., Roberts, D., Smith, M., 1995. Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the brazilian amazon. Remote Sensing of Environment 52, 137–154.

Adam, J., Smith, M., Johnson, P., Gillespie, A., 1986. Spectral mixture modeling: A new analysis of rock and soil types at the viking lander 1 site. Journal of Geophysical Research 91, 8098–8118.

Aguiar, A., Câmara, G., Escada, M., 2007. Spatial statistical analysis of land-use determinants in the brazilian amazonia: Exploring intra-regional heterogeneity. Ecological Modelling 209, 169–188.

Albert, J. M., Mateu, J., Pernias, J. C., 2002. Modelling of spatial point processes derived from a sequence of auto-poisson lattice schemes. Environmental Modelling & Software 17 (2), 105–123.

Alvis, D. S., Pereira, J. L. G., L., D. S. C., Yamaguchi, F., 1999. Characterizing landscape changes in central rondonia using landsat tm imagery. International Journal of Remote Sensing 20, 2877–2882.

Anwar, S., Stein, A., 2012. Detection and spatial analysis of selective logging with geometrically corrected landsat images. International Journal of Remote Sensing 33, 7820–7843.

Anwar, S., Stein, A., 2014a. Spatial-temporal modelling of selective logging: Pattern development over several years. Spatial Statistics.

Anwar, S., Stein, A., 2014b. Use of spatial statistics to investigate early forest degradation activities as detected from satellite images. Spatial Statistics.

Anwar, S., Stein, A., van Genderen, J., 2011. Implementation of the marked Strauss point process model to the epicenters of earthquake aftershocks. Advances in geo-spatial information science, ISPRS Book Series. Taylor & Francis, London, pp. 125–140.

Apan, A. A., A., P. J., 1998. Probing tropical deforestation: The use of gis and statistical analysis of georeferenced data. Applied Geography 18, 137–152.

Asner, G., Keller, M., Pereira, R., Zweede, J., 2002. Remote sensing of selective logging in amazonia – assessing limitations based on detailed field observations, landsat etm+, and textural analysis. Remote Sensing of Environment 80, 483–496.

Asner, G. P., Knapp, D. E., Broadbent, E., Oliveira1, P. J. C., Michael, K., N., S. J., 2005. Selective logging in the brazilian amazon. Science 310, 480–482.

Baddeley, A., 2008. Analysing spatial point patterns in 'R'. Tech. rep., The University of Western Australia, San Jose.

Baddeley, A., Gregory, P., Mateu, J., Stoica, R., Stoyan, D., 2006. Case Studies in Spatial Point Process Modeling. Lecture notes in Statistics 185. Springer, London.

Baddeley, A., Møller, J., Pakes, A. G., September 2008. Properties of residuals for spatial point processes. Annals of the Institute of Statistical Mathematics 60 (3), 627–649.

Baddeley, A., Silverman, B., 1984. A cautionary example on the use of second-order methods for analyzing point patterns. Biometrics 40, 1089–1093.

Baddeley, A., Turner, R., Møller, J., Hazelton, M., 2005. Residual analysis for spatial point processes. Journal of the Royal Statistical Society, Series B 67 (5), 617–666.
URL `http://www.jstor.org/stable/3647612`

Baddeley, A. J., 2010a. Analysing spatial point patterns in r. workshop notes.

Baddeley, A. J., 2010b. Modelling strategies, 371–402.

Baddeley, A. J., Lieshout, M. N. M., 1995. Area-interaction point processes. Annals of Institute of Statistical Mathematics 47, 601–619.

Baddeley, A. J., Møller, J., Waagepetersen, R., 2000. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. Statistica Neerlandica 54, 329–350.

Baker, W. L., 1989. A review of models of landscape change. Landscape Ecology 2, 111–133.

Bateson, A., Curtiss, B., 1996. A method of manual endmember selection and spectral unmixing. Remote Sensing of Environment 55, 229–243.

Boardman, J., 1993. Automated spectral unmixing of AVIRIS data using convex geometry concepts. No. 93 in Summaries, Fourth JPL Airborne Geoscience Workshop. Jet Propulsion Laboratory, pp. 11–14.

Boardman, J. W., Kruse, F. A., Green, R. O., 1995. Mapping target signatures via partial unmixing of AVIRIS data. No. 23 in Summaries, Fifth JPL Airborne Earth Science Workshop. Jet Propulsion Laboratory, pp. 23–26.

Bowman, A. W., Azzalini, A. (Eds.), October 1997. Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations. Oxford University Press, New York.

Broadbent, E. N., Asner, G. P., Keller, M., Knapp, D. E., Oliveira, P. J. C., Silva, J. N., 2008. Forest fragmentation and edge effects from deforestation and selective logging in the brazilian amazon. Biological Conservation 141, 1745–1757.

Brookfield, H., 1999. Environmental damage: distinguishing human from geophysical causes. Environmental Hazards 1, 3–11.

Brown, D. G., Walker, R., Manson, S., Seto, K., 2004. Remote Sensing and Digital Image Processing. Springer, Ch. Modelling land use and land cover change, pp. 395–409.

Comas, C., Mateu, J., 2007. Modelling forest dynamics: A perspective from point process methods. Biometrical Journal 49, 76–196.

Comas, C., Palahi, M., Pukkala, T., Mateu, J., 2009. Characterising forest spatial structure through inhomogeneous second order characteristics. Stochastic Environmental Research and Risk Assessment 23, 387–397.

Contreras-Hermosilla, A., 2000. The underlying causes of forest decline.

Cox, D. R., Isham, V., 1980. Point Processes. Vol. 12 of Monographs on applied probability and statistics. CRC Press.

Cressie, N. A. C., 1993. Statistics for Spatial Data, 2nd Edition. Wiley-Interscience.

Daley, D. J., Vere-Jones, D., 2002. An Introduction to the Theory of Point Processes. Vol. 1. Springer.

Diggle, P. J., June 2003. Statistical Analysis of Spatial Point Patterns, 2nd Edition. A Hodder Arnold Publication.

Fearnside, P. M., 2005. Deforestation in brazilian amazonia: History, rates, and consequences. Conservation Biology 19, 680–688.

Fisher, N. I. (Ed.), January 1996. Statistical Analysis of Circular Data. Cambridge University Press.

Food, organization of the United Nations, A., 2010. Global forest resources assessment 2010.

Gelfand, A. E., Diggle, P. J., Fuentes, M., Guttorp, P. (Eds.), 2010. Handbook of Spatial Statistics. CRC Press.

Gerwing, J. J., 2002. Degradation of forests through logging and fire in the eastern brazilian amazon. Forest Ecological Management 157, 131–141.

Getis, A., Franklin, J., 1987. Second-order neighborhood analysis of mapped point patterns. Ecology 68, 473–477.

Grabarnik, P., Särkkä, A., 2009. Modelling the spatial structure of forest stands by multivariate point processes with hierarchical interactions. Ecological Modelling 220, 1232–1240.

Green, A., Berman, M., Switzer, P., Craig, M., 1998. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Transactions on Geoscience and Remote Sensing 26, 65–74.

Gustafson, E. J., 1998. Quantifying landscape spatial pattern: What is the state of the art? Ecosystems 1, 143–156.

Högmander, H., Särkkä, A., September 1999. Multitype spatial point patterns with hierarchical interactions. Biometrics 55 (4), 1051–1058.

Husen, S., Hardebeck, J. L., 2010. Earthquake location accuracy, community online resource for statistical seismicity analysis. Web site, accessed on 28 July 2014.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., January 2008. Statistical analysis and modelling of spatial point patterns. John Wiley & Sons.

Illian, J. B., Burslem, D. F. R. P., 2007. Contributions of spatial point process modelling to biodiversity theory. Journal de la Soci'ett'e Franÿcaise de Statistique 148, 9–29.

INPE, 2011a. Monitoramento da floresta amazônica brasileira por satélite. Web site, accessed on 10 October 2011.

INPE, 2011b. Mqualidade geométrica das imagens cbers-2 e cbers-2b. Web site, accessed on 10 October 2011.

Isham, V., January 1984. Multitype markov point processes: some approximations. Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 391 (1800), 39–53.
URL http://www.jstor.org/stable/2397529

Jensen, J. R., 2004. Introductory Digital Image Processing – A Remote Sensing Perspective. Prentice Hall Series.

Kagan, Y., Knopoff, L., 1977. Earthquake risk prediction as a stochastic process. Physics of the Earth and Planetery Interiors 14 (2), 97–108.

Kelly, F. P., Ripley, B. D., August 1976. A note on Strauss's model for clustering. Biometrika 63 (2), 357–360.
URL http://www.jstor.org/stable/2335630

kissinger, G., Herold, M., De Sy, V., 2012. Drivers of deforestation and forest degradation: A synthesis report for redd+ policymakers. Tech. rep., Lexeme Consulting, Vancouver, Canada.

Koltunov, A., Ustin, S. L., Asner, G. P., Fung, I., 2009. Selective logging changes forest phenology in the brazilian amazon: Evidence from modis images time series analysis. Remote Sensing of Environment 113, 2431–2440.

Lambin, E. F., 1997. Modelling and monitoring land-cover change processes in tropical regions. Progress in Physical Geography 21 (3), 375–393.

Lambin, E. F., Geist, H., 2006. Land-use and land-cover change. Global Change - The IGBP Series. springer.

Laurance, W., Laurance, S., Delamonica, P., 1998. Tropical forest fragmentation and greenhouse gas emissions. Forest Ecology and Management 110, 173–180.

Law, R., Illian, J., Burslem, D. F. R. P., Gratzer, G., Gunatilleke, C. V. S., Gunatilleke, I. A. U. N., 2009. Ecological information from spatial patterns of plants: Insights from point process theory. Journal of Ecological 97, 616–628.

Li, M., Vere-Jones, D., March 1997. Application of M8 and Lin-Lin algorithms to New Zealand earthquake data. New Zealand Journal of Geology and Geophysics 40 (1), 77–89.
URL `http://dx.doi.org/10.1080/00288306.1997.9514742`

Lorena, R. B., Lambin, E., 2009. The spatial dynamics of deforestation and agent use in the amazon. Applied Geography 29, 171–181.

Lu, D., Batistella, M., Moran, E., Mausel, P., 2004. Application of spectral mixture analysis to amazonian land-use and land-cover classification. International Journal of Remote Sensing 25, 5345–5358.

Lu, D., Moran, E., Batistella, M., 2003. Linear mixture model applied to amazonian vegetation classification. Remote Sensing of Environment 87, 456–469.

Mardia, K. V., Jupp, P. E., November 1999. Directional Statistics. Wiley.

Mateu, J., 2008. Obtaining spatial information using spatial point processes. Web site, accessed on 20 August 2009.

Mateu, J., Usó, J., 1998. The spatial pattern of a forest ecosystem. Ecological Modelling 108, 163–174.

Matricardi, E., Skole, D., Cochrane, M., Pedlowski, M., Chomentowski, W., 2007. Multitemporal assessment of selective logging in the brazilian amazon using landsat data. International Journal of Remote Sensing 28, 63–82.

Matricardi, E., Skole, D., Pedlowski, M., Chomentowski, W., Fernandes, L., 2010. Spatial modelling of deforestation in southern cameroon. Remote Sensing of Environment 114, 1117–1129.

Matricardi, E. A. T., Skole, D. L., Cochrane, M. A., Qi, J., Chomentowski, W., 2005. Monitoring selective logging in tropical evergreen forests using landsat: multitemporal regional analyses in mato grosso, brazil. Earth Interaction 9, 1–24.

Mertens, B., Lambin, E. F., 1997. Spatial modelling of deforestation in southern cameroon. Applied Geography 17, 143–162.

Møller, J., Waagepetersen, 2007. Modern statistics for spatial point processes. Scandinavian Journal of Statistics 34 (4), 643–684.
URL http://ideas.repec.org/a/bla/scjsta/v34y2007i4p643-684.html

Monteiro, A. L., Souza, J. C. M., Barreto, P., 2003. Detection of logging in amazonian transition forests using spectral mixture models. International Journal of Remote Sensing 24, 151–159.

Naranjo, L., 2008. When the earth moved Kashmir. Web site.

nepstad, D. C., Verissimo, A., Alencar, A., Nobre, C., Lime, E., 2005. Markov point processes for modeling of spatial forest patterns in amazonia derived from interferometric height. Remote Sensing of Environement 97, 484–494.

Nuske, R., Sprauer, S., Saborowski, J., 2009. Adapting the pair-correlation function for analysing the spatial distribution of canopy gaps. Forest Ecology and Management 259, 107–116.

Ogata, Y., 1998. Space-time point-process models for earthquake occurrences. Annals of the Institute of Statistical Mathematics 50 (2), 379–402.

Ömer Aydan, 2006. Geological and seismological aspects of kashmir earthquake of october 8, 2005 and a geotechnical evaluation of induced failures of natural and cut slopes. Journal of the School of Marine Science and Technology, Tokai University 4 (1), 25–44.

Perry, G. L. W., Miller, B. P., Enright, N. J., 2006. A comparison of methods for the statistical analysis of spatial point patterns in plant ecology. Plant Ecology 187, 59–82.

Rencz, A., 1999. Manual of Remote Sensing: Remote Sensing for the Earth Sciences. John Wiley and Sons.

Ripley, B. D., 1977. Modelling spatial patterns. Journal of Royal Statistical Society, Series B (Methodological) 39, 172–212.

Roberts, D., Smith, M., Adams, J., 1993. Green vegetation, nonphotosynthetic vegetation, and soils in aviris data. Remote Sensing of Environment 44, 255–269.

Rudel, T., Roper, J., 2002. Forest fragmentation in the humid tropics: A crossâĂŞnational analysis. Singapore Journal of Tropical Geography 18, 99–109.

Santos, J. R., Shimabukuro, Y. E., Duarte, V., Alencastro, G., Paulo, M. L., Silva, P. G., 2003. Linear spectral mixture model as a tool for monitoring deforestation and timber exploitation in the brazilian amazon. In: Proc. SPIE 4879, Remote Sensing for Agriculture, Ecosystems, and Hydrology IV. pp. 320–325.

Schoenberg, F. P., December 2003. Multidimensional residual analysis of point process models for earthquake occurrences. Journal of the American Statistical Association 98 (464), 789–795.
URL `http://www.jstor.org/stable/30045330`

Settle, J., Drake, N., 1993. Linear mixing and the estimation of ground cover proportions. International Journal of Remote Sensing 14, 1159–1177.

Shimabukuro, Y., Smith, J., 1991. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. IEEE Transactions on Geoscience and Remote Sensing 29, 16–20.

Shimatani, K., Kubota, Y., 2004. Spatial analysis for continuously changing point patterns along a gradient and its application to *abies sachalinensis* population. Ecological Modelling 180, 359–369.

Silva, M. P. d. S., Câmara, G., Escada, M. I., 2006. Data mining applications for empowering knowledge societies. Ch. Image mining: detecting deforestation patterns through satellites, pp. 54–74.

Smith, M., Ustin, S., Adams, J., Gillespie, A., 1990. Vegetation in deserts: 1. a regional measure of abundance from multispectral images. Remote Sensing of Environment 31, 1–26.

Souza, C., Barreto, P., 2000. An alternative approach for detecting and monitoring selectively logged forests in the amazon. International Journal of Remote Sensing 21, 173–179.

Souza, C., Firestone, L., Silva, L., Roberts, D., 2003. Mapping forest degradation in the eastern amazon from spot 4 through spectral mixture models. Remote Sensing of Environment 87, 494–506.

Souza Jr., C., Roberts, D., Cochrane, M., 2005. Combining spectral and spatial information to map canopy damage from selective logging and forest fires. Remote Sensing of Environment 98, 329–343.

Stein, A., Bastiaanssen, W., Bruin, S., Cracknell, A., Curran, P., Fabbri, A., Gorte, B., Van Groenigen, J., Van Der Meer, F., Saldona, A., 1998. Integrating spatial statistics and remote sensing. International Journal of Remote Sensing 19, 1793–1814.

Stone, T. A., P., L., 1998. Using multi-temporal satellite data to evaluate selective logging in para, brazil. International Journal of Remote Sensing 19, 2517–2526.

Stoyan, D., Penttinen, A., 2000. Recent applications of point process methods in forestry statistics. Statistical Science 15, 61–78.

Strauss, D. J., August 1975. A model for clustering. Biometrika 62 (2), 467–475.

Turner, B. L., Skole, D., 1995. Land use and land cover change lucc : science research plan. Tech. rep., International Council of Scientific Unions (ICSU), Stockholm.

Turner, R., 2009a. Point patterns of forest fire locations. Environmental and Ecological Statistics 16 (2), 197–223.

Turner, R., 2009b. Point patterns of forest fire locations. Tech. rep., Environmental and Ecological Statistics.

Uhl, C., Verissimo, A., Mattos, M. M., Brandino, Z., Vieira, I. C. G., 1991. Social, economic, and ecological consequences of selective logging in an amazon frontier: the case of tailândia. Forest Ecology and Management 46, 243–273.

Upadhyay, T. P., Solberg, B., 2006. Use of models to analyse land-use changes, forest/soil degradation and carbon sequestration with special reference to himalayan region: A review and analysis. Forest Policy and Economics 9 (4), 349–371.

USGS, 2008. Earthquake hazards program. Web site, accessed on 28 July 2009.

Utsu, T., 1961. Statistical study on the occurrence of aftershocks. The Geophysical Magazine 30, 521–605.

Van Der Meer, F., De Jong, S., 2000. Improving the results of spectral unmixing of landsat thematic mapper imagery by enhancing the orthogonality of end-members. International Journal of Remote Sensing 21, 2781–2797.

Van Lieshout, M. N. M., 2011. A j-function for inhomogeneous point processes. Statistica Neerlandica 65, 183–201.

Vance, C., Iovanna, R., 2006. Analyzing spatial hierarchies in remotely sensed data: Insights from a multilevel model of tropical deforestation. Land Use Policy 23 (3), 226–236.

Verburg, P. H., Schot, P. P., Dijst, M. J., Veldkamp, A., 2004. Land use change modelling: current practice and research priorities. GeoJournal 61, 309–324.

Vere-Jones, D., 1995. Forecasting earthquakes and earthquake risk. International Journal of Forecasting 11, 503–538.

Vere-Jones, D., 2009. Some models and procedures for space–time point processes. Environmental and Ecological Statistics 16, 173–195.

Vere-Jones, D., Li, M., 1997. Application of m8 and lin-lin algorithms to new zealand earthquake data. New Zealand Journal of Geology and Geophysics 40, 77–89.

Vergara, O. R., 2009. Preprocessing of orbital imges at inpe. Web site, accessed on 22 October 2011.

Walker, T. W., 1987. Land use transition and deforestation in developing countries. Geographical Analysis 19, 18–30.

Watt, A. S., 1947. Pattern and process in the plant community. Journal of Ecology 35, 1–22.

Westcott, M., 1972. The probability generating functional. Journal of Australian Mathematical Society 14, 448–466.

Wisner, B., Blaikie, P., Cannon, T., Davis, I., 2003. At Risk: natural hazards, peopleâĂŹs vulnerability and disasters. Vol. 2. Routledge.

Zheng, X.-G., Vere-Jones, D., April 1991. Application of stress release models to historical earthquakes from North China. Pure and Applied Geophysics 135 (4), 559–576.

Zhuang, J., September 2000. Statistical modelling of seismicity patterns before and after the 1990 Oct. 5 Cape Palliser earthquake, New Zealand. New Zealand Journal of Geology and Geophysics 43, 447–460. URL `http://www.informaworld.com/10.1080/00288306.2000.9514901`

Biography

Salma Anwar was born on 01 April, 1977 in Lahore, Pakistan. After obtaining her MSc degree in Statistics from University of the Punjab, she worked voluntaily as a lecturer in APWA college for women, Lahore for two years. In 2005, she qualified as a lecturer in Statistics after passing the Punjab Public Service exam and was appointed as a government lecturer in Statistics at Govt. College for women, Kharain city. After working for two years as a lecturer there, she qualified for the Higher Education Commission of Pakistan (HEC)-sponsored overseas scholarship for MS-leading to PhD in 2007. As a result she joined Department of Earth Observation Science of Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, the Netherlands. In 2009, finishing her thesis with distinction, she received her MSc degree in Geoinformatics. She carried on to do her PhD, working on modeling land use and land cover changes using remote sensing, GIS and spatial statistics. This thesis is the output of her PhD research.

Authors publications

Anwar, S., Stein, A., van Genderen, J. L., 2012. Implementation of the Marked Strauss point process model to the epicenters of earthquake aftershocks *In: Shi W, Goodchild M, Lees B, Leung Y (eds) Advances in geo-spatial information science. Taylor Francis, London* 13 (3), 125-140.

Anwar, S. and Stein, A., 2012. Detection and spatial analysis of selective logging with geometrically corrected Landsat images. *International Journal of Remote Sensing* 33, 7820-7843.

Anwar, S., Stein, A. (Accepted with minor revision). Use of spatial statistics to investigate early forest degradation activities as detected from satellite images. *Spatial Statistics*.

Anwar, S., Stein, A., 2014. Spatial-temporal modelling of selective logging: Pattern development over several years. *Submitted to Spatial Statistics*

ITC dissertations

A complete list of ITC dissertations is online on the ITC website: www.itc.nl/research/phd/phd_graduates.aspx.

This dissertation has number 257.