Université catholique de Louvain
Faculté d'ingénierie biologique,
agronomique et environnementale

Modeling the Forest Cover Loss in Democratic Republic of the Congo

Methods Development and Application at National Scale

Kibambe Lubamba Jean-Paul

JUILLET 2013

Thèse présentée en vue de l'obtention
du grade de docteur en sciences agronomiques
et ingéniérie biologique

Faculté d'ingénierie biologique,
agronomique et environnementale

Université catholique de Louvain

Composition du jury

Président:
Pr Jacques Mahillon (UCL, Belgique)

Promoteur:
Pr Pierre Defourny (UCL, Belgique)

Lecteurs:
Pr Christopher O. Justice (UMD, USA)
Pr Raymond Lumbuenamo Sinsi (UNIKIN, RDC)
Pr Patrick Bogaert (UCL, Belgique)
Pr Quentin Ponette (UCL, Belgique)

Usipoziba ufa utajenga ukuta.

If you don't seal a crack (in the wall),
you will (have to) (re)build a (complete) wall.

Proverbe swahili

Acknowledgments

Cette thèse a été pour moi une opportunité extraordinaire de travailler avec des personnes exceptionnelles à de nombreux points de vue. Au nombre de ces personnes figure le Professeur Pierre Defourny, sans qui cette thèse n'aurait jamais commencé. Pierre a su me guider avec dextérité, patience, exigeance et bonne humeur tout au long de ce travail de recherche, malgré son emploi toujours surchargé. J'ai été particulièrement frappé par sa capacité à s'intéresser à la fois aux questions de recherche ainsi qu'aux menus détails de ma passion pour le football... Je voudrais par ces mots lui témoigner toute ma gratitude pour sa disponibilité, son écoute attentive et son enthousiasme contagieux.

Professor Chris Justice is also one of these people. I remember when we conducted field work together in Basankusu, we were searching for primary humid tropical forest blocks in the Equateur Province. During this trip, Chris suggested I conduct an internship in his laboratory at the University of Maryland's Department of Geographical Sciences, to develop a spatially-explicit model of land-use and land-cover cover change for the Maringa-Lopori-Wamba Landscape. This was my first visit to the U.S. and a tremendous incentive for research at the beginning of this thesis. Chris did not hesitate when asked to be a part of this jury, and he traveled from the U.S. to Belgium twice this year to attend both the preliminary defense and the public presentation, despite his overwhelming agenda. Thank you, Chris, for your support and your contribution to the achievement of this thesis..

J'ai eu l'immense privilège d'être étudiant du Professeur Raymond Lumbuenamo à l'Université de Kinshasa. Dès ma toute première séance de cours, j'ai été convaincu que celui qui est devenu quelques années plus tard l'un des initiateurs du Département de Gestion des Ressources Naturelles au sein de la Faculté des Sciences Agronomiques, allait m'ouvrir à de nouveaux horizons. Un grand merci à toi, Prof Raymond !, pour ton incitation permanente à nous surpasser malgré nos conditions de travail difficiles à l'Université de Kinshasa. Merci aussi pour ta confiance en me laissant travailler à tes côtés au sein du laboratoire de cartographie numérique et de télédétection de l'Observatoire Sattelital des Forêts d'Afrique Centrale. La réalisation de cette thèse a été pour moi une manière de te rendre un tout petit peu de ce que tu m'as transmis, même si je suis pleinement convaincu d'être encore loin du compte !

Le Professeur Patrick Bogaert n'a pas eu de cesse à me pousser à plus de rigueur dans l'interprétation des résultats de cette thèse, dans l'implémentation des tests statistisques adéquats, ainsi que dans le développement des algorithmes. Patrick, je voudrais te remercier pour tout cela, et pour ton légendaire sens de l'humour qui a toujours égayé nos conversations à la cafétéria ENGE. Mes remerciements s'adressent aussi au Professeur Quentin Ponette pour sa disponibilité à faire partie de ce jury de thèse, sans doute dicté par son enthousiasme à contribuer à la formation des futurs cadres de l'administration forestière congolaise, au travers de son implication dans les projets de formation à l'Université de Kisangani. Je voudrais aussi remercier le Professeur Jacques Mahillon d'avoir bien voulu présider la séance de la soutenance publique.

Mes remerciements vont ensuite tout naturellement à mes parents, que j'aime nommer par leurs noms *in extenso*: Kibambe Mikombe Ngoy Grégoire et Ngaj-Kat-A-Mwinkew Jacqueline. Papa et Maman, vous m'avez inculqué dès mon plus jeune âge l'amour du travail bien fait et l'abnégation face aux anicroches de la vie. Merci à tous les deux pour votre dévouement inlassable à m'apporter votre soutien, vos encouragements, et votre générosité. Que mes frères et soeurs Serge, Da Claude, Ya Tity, Ya Nina, et Max (le petit Docta - Heureux mariage dans quelques semaines !) trouvent ici l'expression de mes sentiments les plus dévoués à leur égard. Je voudrais aussi dire à mes neveux Nadine, Paty, JR, Go; Laeticia, Greg et Maeva que votre Tonton "le plus sympa" pense toujours très fort à vous. J'ai en ce moment une tendre pensée pour Da Brigitte qui nous a quittée bien trop tôt, alors que nous avions encore tant besoin d'elle.

Papa Etienne Kalasa Munongo (qui nous a aussi quitté alors qu'il avait encore tellement à nous apprendre de la vie) et Maman Gaby ont depuis ma tendre enfance été mes seconds parents, pour devenir par la suite mes beaux-parents... Ils ont tous les deux largemment contribué à ce que cette thèse voit le jour, ne serait-ce que par leur présence aimante et exigeante à la fois pour souder les liens de la grande famille... Et elle est tellement grande que je ne saurai citer tout le monde nommément... Mais je voudrais faire ici un clin d'oeil à Tonton Godé (et sa formidable famille dans le Connecticut où je me suis toujours senti chez moi), à Tonton Vital "le Sénateur" et Da Manena (ainsi que leurs charmants "cocos" Anaïs, Dadou, Gabou et Manunette), à mon parrain Papa Alain et ma marraine Da Lydie (ainsi que mes autres "cocos" Lyn, Dan, et Sam-Sam!), à Tonton Gilles et Da Bwile (oui, je dois vous rendre visite dans le Michigan, je sais !!!) ainsi que le "garde du corps" et ses soeurs, à Tantine Anto, à Marie-B. (MB !!!), à Tonton Steve pour son immense disponibilité ainsi qu'à toute "l'équipe choc" dont notamment Junior, Kali et Pinya (le pire petit ya Vieux na ye !), à Tanya et Christy (Naine !!!) ainsi qu'à Denis et Magaly et leurs deux merveilleux enfants Teteka et Fafaël mon filleul. Je ne peux terminer cette évocation sans remercier tout particulièrement Tonton Tshaki et Tantine Christy de m'avoir toujours porté dans leur coeur... Et avec eux, Isabelle et son bout de chou, Ronan et Carole sans oublier Youna et le boss Hugo.

Un grand merci à l'ADRI pour la bourse de doctorat qui m'a été accordée eet qui m'a donnée l'immense chance de travailleur avec de collègues d'une rare compétence et d'une bonne humeur intarrissable. Special thanks to Janet for her contribution to this thesis. Since more than four years, we are working together on DRC spatial modeling, making the collaboration between UCL and UMD more effective. Il m'est extrêmement difficile de rappeler ici, comme je voudrais tant le faire, la spécificité de chacun de mes collègues... Merci à Julien pour ses conseils toujours pertinents pour exécuter des scripts Python et pour sa combativité sur les terrains de sport, à Sophie pour (entre autre) sa disponibilité et son travail remarquable au sein du labo, à Astrid qui termine sa thèse avec moi pour les nombreux moments passés à discuter de nos thèses respectives et la promesse de se retrouver autour de quelques triple "K" pour fêter cet aboutissement comme il se doit, à Inès toujours partante pour aggrémenter la vie du labo, à Aline à qui on doit notamment notre nouvelle salle Quételet, à Doriane, à Céline (tu sais bien que tu es toujours la bienvenue dans le "paysager du couchant"), aux deux Catherine, à Benjamin, Damien (DJ !!!), à François, à Thomas, à Corentin, à Sarah pour ses précieux conseils en à tats, à Jolan et à Emmanuel pour ses idées novatrices comme celle de prendre un (au moins un) verre à chaque fois que le tableau des publi ENGE est rempli... Un merci tout particulier à Raf pour son enthousiasme communicatif et à Franz Wald TEX (tout est dit !). Y Manu, pensaste que te habia olvidado o qué? Nunca, Tio... ahora si, puedes llamarme Doctor ;-)! Mention toute particulière à mes collègues de paysager Pierre, Maïté et l'inépuisable et insubmersible Quentin qui a tellement contribué à la bonne implémentation du modèle de déforestation !!! Je ne peux oublier mes anciens collègues de labo Greg, Eduardo, le général Peeters, Cozmin, Baudouin, Carlos et le "congolais" Xavier! Enfin, un tout grand merci à Pascale sans qui le bon déroulement de cette thèse aurait été impossible. Son soin particulier dans la gestion de mes dossiers administratifs a toujours été remarquable !

Mes collègues et amis congolais sont à la fois nombreux, précieux et inoubliables. Merci à toi Paul pour ton exemple, à Yves-Daddy, à Papy ("le Formateur"), à Mukulukulu, au Président David, à Edouard et Fanny (une pensée pour la petite), à Didier ("le Ministre de la Santé"), au Ministre Martin (et toute sa famille), à Patrick, Eddy, Yolande, Confiance, Dibock, Maximme, Dodo, Martine, Florence, André... à Marcelline et à l'homme de la Lomako,... Une mention toute particulière à Link Burrows, mon frère et ami, à son épouse Nancy et à leur (ma) fille Gracia.

Je ne peux terminer sans mentionner Didier Devers avec qui j'ai travaillé de longues années à l'OSFAC et à l'OFAC. Tu es devenu au fil des années un ami précieux. Merci pour tout, à toi et à Ale. Je ne peux non plus oublier Tonton JR, pour son soutien, son appui, sa présence,... Je n'oublierai jamais l'agréable séjour chez toi à Brookings, SD ainsi que l'opportunité que tu m'as donné de pouvoir utiliser pour la première fois de ma vie ArcView 3.2 sur ton ordinateur à Salongo... Je n'aurai jamais imaginé faire de la géomatique si je ne t'avais pas rencontré.

Pour finir, je voudrais dire à la femme de ma vie, Vivi, que ton soutien permanent dans les moments de travail intense ont été ma secrète motivation de continuer à aller de l'avant. Ce document n'aurait jamais vu le jour sans toi, que je connais depuis mes 8 ans... Tu m'as, en plus, fait le plus beau des cadeaux : mon adorable Anaëlle (Akilu !!!), Choupinette, Choupinenette, Choupi ! Vous êtes les repères de ma vie... Je vous aime tellement !

Contents

List of Figures

List of Tables

List of Acronyms

ABM	Agent-Based Model.
BAU	Business-As-Usual.
CA	Cellular Automata.
CARPE	Central Africa Regional Program for the Environment.
CHANS	Coupled Human And Natural Systems.
CIA	Central Intelligence Agency.
COP	Conference of the Parties.
DCRM	Double-Constrained-Regression Method.
DRC	Democratic Republic of the Congo.
EMU	Elementary Mapping Unit.
EO	Earth Observation.
EPU	Elementary Processing Unit.
ETM+	Enhanced Thematic Mapper Plus.
FACET	Forêts d'Afrique Centrale Evaluées par Télédétection.
FAO	The Food and Agriculture Organization of the United Nations.
FRA	Forest Resources Assessment.
GPS	Global Positioning System.
GPW	Gridded Population of the World.
HM	High Mosaic.
IAM	Integrated assessment model.
IGPB	International Geosphere-Biosphere Programme.
IHDP	International Human Dimension Program.

INCR INCRement.

ISSC International Social Science Council.

LCCS Land-Cover Classification System.

LCLCC Land-Cover and Land-Cover Change.

LCLUC Land-Cover and Land-Use Change.

LM Low Mosaic.

LR Local Road.

MLPNN Multi-Layer Perceptron Neural Network.

MMU Minimum Mapping Unit.

NASA National Aeronautics and Space Administration.

NDVI Normalized Difference Vegetation Index.

NR National Road.

NTC No Tree Cover.

OB Object-Based.

OFAC *Observatoire des Forêts d'Afrique Centrale*.

OSFAC *Observatoire Satellital des Forêts d'Afrique Centrale*.

PA Protected Area.

PB Pixel-Based.

PF Primary Forest.

PR Prioritary Road.

REDD+ Reducing Emissions from Deforestation and Forest Degradation initiative.

RGC *Référentiel Géographique Commun*.

SF Secondary Forest.

SLC Scan Line Corrector.

SLU Smallest Legend Unit.

SRTM Shuttle Radar Topography Mission.

SSE Sum of Squared Errors.

TC Tree Cover.

TM Thematic Mapper.

UCL Université catholique de Louvain.

UNFPA United Nations Population Fund.

UNPD United Nations Population Division.

WO Woodlands.

WRI World Resources Institute.

WT Water.

Introduction

Earth surface has been continuously transformed by human activities through multiple and complex interactions involving social, bio-geophysical, economic, and cultural factors that operate at micro-, meso-, and macro-scales. These interactions include among others the variations in the CO_2 fluxes between the Earth surface and the atmosphere, the modifications in food and fiber production, the changes in both the ecosystem functioning and land-use and -cover (Haberl *et al.*, 2001a,b; Lambin *et al.*, 2003; Gallopin, 2006; Mooney *et al.*, 2009; Rounsevell *et al.*, 2012). The analysis of land-use and land-cover change (LCLUC) is multi-disciplinary and multi-scalar per nature (Jepson and Millington, 2008; Rounsevell *et al.*, 2012) and seeks to understand the prevailing human and environmental factors.

In our contemporaneous era, Earth surface changes have reached unprecedented thresholds (Rockström *et al.*, 2009) observation data particularly allow the monitoring of changes in natural systems, which are analyzed in conjunction with human activities in the coupled human and natural systems (CHANS) research area (Turner *et al.*, 2004). However, the complex interactions between these two systems have not been well understood yet, which pledges for developing (integrated) models of land-use and -cover change that can contribute to unravel these complexities (Liu *et al.*, 2007).

The coupled human and natural systems

The CHANS aims at understanding, documenting, and explaining the human-environment interactions through the monitoring of land-use and -cover (LULC) dynamics, as well as spatially-explicit modeling of the environmental and societal systems (Rindfuss *et al.*, 2004; Turner *et al.*, 2007).

The inherent complexity of the CHANS requires a holistic approach. This has to handle cross-scale interactions and incorporate in the same framework the dynamic of land systems, the consequences of the human-induced land system changes, and the integrative land system analysis and modeling. The general scope is both a sustainable use of land resources and the implementation of appropriate policies for the land management (Riebsame *et al.*, 1994; Liu *et al.*, 2007; Stannard and Aspinall, 2011).

Integrated assessment models

Integrated assessment models (IAM) play an important role in understanding and quantifying the Earth system, and specifically the impacts of land-use and land-cover change (Hibbard *et al.*, 2010). They are indeed focused on improving the description of land-use and -cover change dynamics (Piao *et al.*, 2007). Satellite data, in conjunction with field measurements and historical reconstructions, are useful tools for exploring past, present, and future land-use/cover (Pongratz *et al.*, 2008; Sertel *et al.*, 2010). A point of particular interest is therefore linking LULC dynamics to their range of potential drivers, which can be achieved through spatially-explicit modeling.

Several studies have used spatial models to link socio-economic data, human behaviors, land management policies, and remote sensed information. Examples of these include the impact of stocking practices by farmers-pastoralists on the vegetation cover (biomass) in arid-lands through the analysis of the Normalized Difference Vegetation Index (NDVI) (Archer, 2004); the emergence of *hybrid* land-covers — that were mapped through satellite images analysis in combination with information about the household agricultural production — resulting from particular land management policies (Robbins, 2001); the impact assessment of the human settlement distribution (dense/sparse) on the efficiency of established protected areas (Liu *et al.*, 1999, 2004); or the impact of the regulatory policies related to small scale and intentionally set fires — that can be tracked using satellite images (Davies *et al.*, 2009) — on the human livelihoods (Laris, 2002).

Integrating the human dimension into the analysis of LULC dynamics remains a challenging task. Indeed, the required data are most likely to be collected from variant (and often not coordinated) initiatives that can be either research or nature conservation projects, or government-supported surveys. This ultimately leads to a heterogeneous set of information, which hampers their further comparison/use in the framework of an integrated analysis. This was corroborated by Janetos (2004) who observed that the direct (spatial-scale) matching between these types of data — such as previously found for a limited number of land-cover change case studies — was likely to be fortuitous.

Today, the growing awareness about climate change issues (re-)boosted the use of integrated assessment models, which tend to bridge the gap between the observation, measurement, modeling and decision-making communities to generate useful information for policy making (Warren *et al.*, 2008; Hibbard *et al.*, 2010).

Land-use and -cover change models

Land-use and land-cover change has emerged such as one of the key independent and cross-cutting themes among several major international initiatives including the International Geosphere-Biosphere Programme (IGPB), the International Social Science Council (ISSC), the International Human Dimension Program (IHDP), and the NASA's LCLUC research program (Moran *et al.*, 2004; Rounsevell *et al.*, 2012). Furthermore, changes in both land-cover and land-use are so intimately linked that they have to be investigated through the coupled human-environment system analysis that is the core concept in the majority of IAM. These models intend to understand the land-use/cover change dynamics, develop hypotheses that can be tested empirically, and investigate future land-use/cover changes and/or evaluate scenarios for use in assessment activities (Brown *et al.*, 2004).

Drivers of land-use/cover change are generally integrated into empirically-fitted or dynamic spatial models. The first type relies on a statistically matching between the LCLUC and a set of (spatial) predictors. These latter mostly include the transportation network proximity, the demography, as well as biophysical factors such as the slope and the soil types. Fitted models commonly use logistic regression functions that describes either the probability occurrence of a particular land category, or the location transitioning from one land category to another (Chomitz and Gray, 1996; Lambin, 1997). Given, however, the fact that log-linear relationships between the dependent variables and the predictors can be restrictive, these linear assumptions are often relaxed using non-linear fitted models (Brown *et al.*, 2002). Also, fitted models typically assume a temporal stationarity in the relationship, which can be tackled by using econometric approaches such as those developed for analyzing time series or panel data (Hsiao, 2006). Spatial autocorrelation, that is a consequence of the diffusional nature of many land change processes (Walker *et al.*, 2000), has also to be addressed since it can lead to fallacious results about the relationships between the land-use/cover changes and the selected predictors. The endogeneity that could exist between the land-cover changes and some of the predictors (Chomitz and Gray, 1996) has also to be taken into account in order to avoid an artificially raised explanatory power of the model.

Dynamic process models applied to LCLUC seek, for their part, to mimic the most important interactions that are found in the CHANS. They typically rely on simulations that constitute their most central emphasis. Though, simulations can also be computed using fitted models (Brown *et al.*, 2002). Dynamic agent-based and cellular automata (CA) models are part of dynamic process models.

Agent-based models (ABM) give an explicit attention to the interactions between agents of land change. These models can therefore simulate the (emergent) properties of systems that are both macro-scale and not predictable from the observation of individual units of simulation. ABM typically use a *bottom-up* approach for exploring the land-use/-cover changes (Bousquet *et al.*, 1998; Polhill *et al.*, 2001; Verburg *et al.*, 2006). However, these types of models most often require large amount of field-based data.

CA models are typically based on a top-down approach and use a grid of cells that can represent parcels of land. Each cell has its own characteristics while its changing status depends on its state and the state of its neighbors. The collective behavior of cells that influence each other can lead to emerging patterns, so that these models are expected to capture important dynamics (e.g., road-driven and spatially random changes) of the coupled human-environment. One of the most challenging issues for improving LCLUC modeling based on CA, apart from the need to ameliorate the model calibration and validation, seems to be the incorporation of heterogeneity and dynamism within the basic rules of neighborhood interactions. Examples of these improvements are the integration of the attraction and repulsion effects between the diverse land-cover types, within the neighborhood of each cell that are modeled such as CA function states (Barredo *et al.*, 2003). The integration of functions such as those expanding existing patterns or generating new patches (Soares-Filho *et al.*, 2002) also contributes to introduce the required dynamism and heterogeneity into the CA transition rules.

Coupling both CA and ABM models allows in particular integrating the non-spatial properties derived from ABM models into geospatial simulations (Castle and Crooks, 2006; Bithell and Brasington, 2009).

Overall, spatial models of land-use/cover changes intend to unravel the complex interactions involved in the modeled system by using artificial representations, so that they are able to provide useful information for the land management and planning. These models can therefore play a mayor role in undertaking a structured analysis so that future possible evolutions of land can be explored through *'what-if'* scenarios.

Land-use and -cover change scenarios

Land-use planners have been using scenarios of land evolutions since several decades (Xiang and Clarke, 2003). Spatial modeling of land-use/cover changes typically uses a set of scenarios to simulate variations in the drivers of changes, so as future changes can be predicted.

The different scenarios are, however, contingent on the uncertainty about the future. This is mainly a consequence of the possible emergence of novel interactions in the modeled system. This uncertainty is at the same time a constraining factor and a valuable opportunity for land-use/cover change modeling since although the future is unknown, it can be created or anticipated by the plans and the action of people. The scenario has in this sense a dual function that is a scientific inquiry about the future and the real world planning for the future (Xiang and Clarke, 2003).

The scenario planning method emerged in large part from the necessity to deal with the uncertainty in the scenarios. The rationale is to consider a variety of possible futures that include many of the important uncertainties in the system, rather than to focus on the accurate prediction of a single outcome (Peterson *et al.*, 2003). A set of alternative land evolutions that are typically shaped by the current dynamics and future possible events are thus analyzed. The few alternatives retained are converted into dynamic stories (*storylines*) by adding a credible series of external forces and actors' responses.

Building scenarios is a participative exercise *per se* that involves policy makers, land managers, scientists, and other stakeholders (Swetnam *et al.*, 2011). One of its main goals is examining the future in light of a specific question, also termed such as the 'big picture' (e.g., what will be the major trends in the forest cover over the next 30 years?), which is typically raised by (inter) national policy makers (Xiang and Clarke, 2003) . The 'big picture' starting point helps distinguishing the aspects of the system that are likely to be known in the future from those that are not, which contributes to assess the uncertainty around the possible evolutions of land-cover dynamics.

The uncertainty about future land-use/cover dynamics is exacerbated in sparse data contexts. This is the case for the major part of developing countries where both the past and the actual conditions/dynamics of land-use are poorly known. In these countries, only few land-cover change data are often available, which can hinder building scenarios about the possible future changes. At the same time, land-cover/use scenarios are critically needed in these countries since they are more vulnerable to the adverse effects of the global environment change. Spatial modeling of future land-use/cover changes based on simplified scenarios is sometimes the only viable way to assess these adverse effects.

Spatial modeling in sparse data contexts

Remote sensing-based land-cover data are the unique viable source of information for land-use/cover change modeling in regions where field-based measurements are of poor quality. In these regions, socio-economic data are rarely at appropriate spatial and temporal scales in regards to the remote sensed information (Rindfuss *et al.*, 2004). Remote sensing, in contrast to socio-economic data, can indeed provide both a continuous coverage and repetitive observations. In order to match the temporal resolutions of both data sources, socio-economic data are commonly inferred through (back) projections that little account for the methodology used at the time the native data sets were produced.

In addition, available socio-economic data are often summarized at sub-national levels (Janetos, 2004). In most cases, these data need to be disaggregated at the pixel level prior their integration into spatial models. Yet, this disaggregation is based on proxy-information that often lack in developing countries. Moreover, the various existing disaggregation methods need to be carefully evaluated since the relationships between the pixel-level information and the summarized data may be scale-dependent (Janetos, 2004; Rindfuss *et al.*, 2004; Dendoncker *et al.*, 2006).

Another concern in, although not restricted to, sparse data contexts is the necessary validation of existing data prior to any land-use/cover change modeling (Townshend *et al.*, 2004). The standards referring to both the methodology that led to the data production and the appropriateness of data in terms of accuracy, scale, and time-periods are also common problems found in scarce data contexts. These concerns are of paramount importance for countries involved in international initiatives (e.g., the Reducing Emissions from Deforestation and Degradation program in developing countries) requiring the assessment of country-based data by third parties, prior to any financial allocation rewarding the avoided deforestation/degradation. For this assessment, data have to meet some level of international standards in terms of precision, transparency, repeatability, and uncertainty evaluation.

On the other hand, global data are becoming increasingly available. These include among others the Food and Agriculture Organization of the United Nations (FAO) database[1], global land-cover data sets[2], global population data sets[3][4], global road data[5], etc. These however are still of low precision in developing regions since they are most often built on country-based data. As a consequence, diverse data collection efforts are currently funded in developing countries (Awusabo-Asare *et al.*, 2009). While being highly commendable, these government- and non-government-subsidized efforts which are often undertaken in crisis contexts, have often restricted financial resources, and therefore seldom gather information that are valid at the national scale. Such a data collection ultimately contributes to a proliferation of various information that are not straightforwardly comparable although they refer to both the same topic and region.

The Democratic Republic of the Congo: a poor data environment

The Democratic Republic of the Congo (DRC) is slightly over 2 millions km^2, almost equivalent to Western Europe. It is the biggest country in sub-saharan Africa and the most populated of central Africa, with about 55,000,000 inhabitants in 2005 (De Saint Moulin, 2006). The country is administratively subdivided into different levels. From the smallest to the largest, the administrative units are: the 'Village' level, the 'Chefferie' level, the 'Groupment' level, the 'Sector' level, the 'Territory' level, the 'District' level, and the 'Province'.

The most recent population census was conducted in 1984, while the United Nations Population Fund (UNFPA) strongly advocates that population and

[1] `http://faostat.fao.org/?lang=en` [Accessed 12 January 2013]
[2] `http://due.esrin.esa.int/globcover/` [Accessed 12 January 2013]
[3] `http://sedac.ciesin.columbia.edu/data/collection/gpw-v3` [Accessed 12 January 2013]
[4] `http://www.ornl.gov/sci/landscan/` [Accessed 12 January 2013]
[5] `http://www.ciesin.columbia.edu/confluence/display/roads/` [Accessed 12 January 2013]

housing censuses should be taken at least every ten years [6]. Population data from this census were aggregated to the Territory administrative level, which corresponds to areas that can cover about 50,000 km^2. From the 1984 census, the DRC total population was estimated to be about 30,731,000 inhabitants (Ngondo *et al.*, 1992). All currently available spatial data on DRC's population, including the Gridded Population of the World (Deichmann *et al.*, 2001; Balk and Yetman, 2004), LandScan (Dobson *et al.*, 2000), Afripop (Tatem *et al.*, 2007), were computed based on the aggregated 1984 census data projections.

Due to several decades of mismanagement and the persistent political unrest, national institutions in charge of collecting and compiling socio-economic and geographical country-data only provide fragmentary information. This contributes to a pervasive burgeoning of field data collection efforts (Blaes, 2009), which deliver (spatial) data that most often have divergent precision, and can be of poor quality or incomplete.

The DRC economy — dominated by the primary sector (i.e., agriculture, forest, and other extractive industries) — is largely informal (Debroux *et al.*, 2007). Agriculture is mainly subsistence-oriented and the vast majority of the DRC's households rely on the forest products for energy, food, and income (Debroux *et al.*, 2007; Tollens, 2010; Schure *et al.*, 2011). Besides, agriculture is also the sector that offers the best perspectives for a sustained economic growth, which would benefit to the majority of the population (Tollens, 2003).

The DRC hosts approximately 50% of Congo Basin dense forests (Verhegghen *et al.*, 2012), the world's second largest tropical forest. It is also of high conservation priority given that about 50% of priority areas termed as 'Lansdcapes' within the Central Africa Regional Program for the Environment (CARPE) initiative are located within DRC boundaries (Vande Weghe and Devers, 2006). At the same time, the DRC has been identified such as a UN-REDD (United Nations Reducing Emissions from Deforestation and Forest Degradation initiative) leader country. In fact, it has been widely recognized during the 17th Conference of the Parties (COP-17) in Durban that DRC made significant efforts (including strong political commitment as well as implementation of operational forest monitoring tools) towards the implementation of REDD+ activities in the country. The DRC consequently benefits from an important worldwide attention resulting in a growing demand for the monitoring of forest cover and changes (Tadoum *et al.*, 2012). In late 2011, the DRC launched a new national forest monitoring system, built on the Brazilian 'wall-to-wall' deforestation mapping initiative[7]. This government-supported initiative can be seen as evidence of political commitment in the DRC towards operational forest monitoring systems using satellite images (Brady *et al.*, 2010).

Forest cover monitoring in the DRC relies almost exclusively on remote

[6] `www.unfpa.org/pds/census.html` [Accessed 12 January 2013]
[7] `http://www.rdc-snsf.org/` [Accessed 15 january 2013.]

sensing. The lack of both infrastructure and financial resources, coupled with safety issues prevented performing nationwide scale field-based measurements. Though, there were (are) numerous forest data collection efforts at local scales (Bogaert *et al.*, 2008), these were (are) not designed to provide consistent information for the all country.

The drivers of forest change in DRC have been recently identified by Defourny *et al.* (2011). These include in particular population growth, the proximity to human settlement, agricultural expansion, and protected areas. However, the necessary data that can allow quantifying these causes are generally of poor quality, which hampers the forest cover change modeling (Ludeke *et al.*, 1990; Apan and Peterson, 1998; Lambin *et al.*, 2000; Nelson and Geoghegan, 2002). Additionally, the forest monitoring in the DRC typically results from various initiatives, which deliver different estimates that are sometimes not straightforwardly comparable since they are based on various methodologies (Potapov *et al.*, 2012; Ernst *et al.*, 2013). The rate of gross forest cover changes in DRC is relatively low by comparison to other high forested countries such as Brazil. For the period 1990-2000, the FAO Remote Sensing Survey of the Global Forest Resources Assessment (FAO-FRA) estimated the annual rate of forest change in DRC to be about 0.4% (FAO, 2001). For the same period, Ernst *et al.* (2013) derived annual rates of gross deforestation (0.15% $\pm$ 0.02%), gross forest degradation (0.07% $\pm$ 0.01%), gross reforestation (0.04% $\pm$ 0.01%), and gross forest regeneration (0.02% $\pm$ 0.00%). Between 2000-2005, these rates increased to 0.32% $\pm$ 0.05%, 0.16% $\pm$ 0.03%, 0.10% $\pm$ 0.03%, and 0.04 $\pm$ 0.02 % respectively for the four processes (Ernst *et al.*, 2013), while the annual rate of forest change was estimated at 0.2% according to FAO (2006). Besides, Potapov *et al.* (2012) estimated the gross forest cover loss to be about 0.22% *per annum* for the period 2000-2005, which increased to 0.25% between 2005-2010. While these rates are also the highest found among central African countries (Ernst *et al.*, 2013), little has been done to simulate possible future forest cover changes in DRC at the national scale.

Scope and objectives

The overall objective of this thesis is to investigate the possible futures of the DRC forests by developing a spatially-explicit and dynamic model of forest cover loss at the national scale. The main driving forces of these losses are translated into appropriate spatial variables that include specifically the distribution of the human population, the accessibility to major towns and big villages, the proximity to existing secondary forests, the location of protected areas, and the most reliable estimates of forest cover change.

To fulfill the goal of the research, two main assumptions are made and addressed through three specific research questions. Firstly, it was hypothesized that estimates of forest cover change from most recent remote sensing-based products for the DRC contain sufficient information to characterize the forest cover dynamics. These estimates are, however, not straightforwardly comparable since they are derived from independent methods that are built on their own specific assumptions. A detailed analysis of these products is therefore required. Secondly, the human population distribution and the accessibility to major towns were assumed to be the main predictors of both the amount and spatial patterns of forest cover change. Retrieving these variables in the poor data context of the DRC requires, however, the implementation of appropriate spatial models. In fact, population data are only available at coarse scale and the poor condition of the existing road network leads to using alternative transportation systems. Further simulations of forest cover change have therefore to clearly show how these two variables in particular are accounted for to derive both the amount and the patterns of change.

The first research question is formulated as follows:

1. *Can the results from independent remote sensing methods (i.e., pixel- and object-based classifications) for monitoring forest cover changes be combined in order to derive consolidated estimates?*

Pixel- and object-based image classifications led to diverse forest cover mapping initiatives throughout the tropical countries (Duveiller *et al.*, 2008; Hansen *et al.*, 2008a; Potapov *et al.*, 2012; Ernst *et al.*, 2013). The products obtained from these two approaches can yield different results that are not straightforwardly comparable since their underlying assumptions are different. For many Congo Basin countries (Ernst *et al.*, 2013), the necessary comparison of these products is still to be achieved for a proper assessment of the past and current forest change trends. This comparison also contributes to meet the standards set in the framework of international initiatives such as the REDD+, which incentivizes developing countries to provide transparent, replicable, and consistent estimates of forest cover change.

The first specific objective of this research seeks to develop a dedicated methodology to compare the latest estimates of forest cover change in DRC. These are based on pixel- and object-based image processing using Landsat images. The developed comparison analysis allows highlighting and discussing the divergences to further derive consolidated annual rates of four different forest cover change processes that are the deforestation, the forest degradation, the forest regeneration, and the reforestation (Chapter 1). The comparison method is thought to be applicable to other Congo Basin countries facing similar issues since the monitoring of the tropical forest changes is in the agenda of several initiatives that attempt to attenuate their impacts on the local, regional, and global climate. It is therefore of high importance to provide policy makers with consolidated estimates of rate of change, particularly for high forested and least developed countries that lack sufficient resources to conduct the monitoring of their forested land on a regular basis.

The second research question is related to the driving forces of forest cover change in the DRC. These are most often not properly quantified using spatial variables, for their further use into spatially-explicit models of forest cover change. Indeed, the proximity existing network is largely computed simply as distance to roads (Chomitz and Gray, 1996; Mertens and Lambin, 2000; Nelson, 2008; Nackoney and Williams, 2012), while the human population distribution is mostly obtained from global population data sets which have a lower performance for many developing countries (Baker *et al.*, 2010). This thesis thus addresses the following question:

2. *Can spatial variables quantifying the main driving forces of the forest cover loss be appropriately retrieved in a poor data context of the DRC?*

The forest cover loss in DRC is correlated to population density and the proximity to human settlements, which can be approximated using respectively the human population distribution and the accessibility to major towns and big villages (Defourny *et al.*, 2011; Ernst *et al.*, 2013). The latter is contingent on the performance of the transportation system that link rural to urban areas. However, very little attention has been paid to develop a methodology that translates how people — in addition to the use of the existing road network — make use of navigable rivers in conjunction with *off-road* travel for accessing to the forest resources. These alternative transportation systems are traditional and are commonly used in the absence of adequate road networks.

Besides, DRC major cities generally have a poor supply of electricity, leading most of urban populations to rely on fuelwood collection and charcoal production in *surrounding* rural areas to secure domestic energy needs (Tollens, 2010; Schure *et al.*, 2011). The DRC urban population represents about 30% of the total population and the remaining is unevenly distributed (De Saint Moulin, 2006). However, available population distribution data for the DRC are generally of poor quality. Census data that serve as the basis to infer on the spatial distribution of the human population are outdated or only available at coarse scales. In addition, developing countries often lack the proper resources to support large scale censuses (Baker *et al.*, 2010). As a consequence, although global population distribution data are becoming increasingly available, the corresponding information for developing countries — such as the DRC — are still of lower reliability (Linard *et al.*, 2012).

The second specific objective tackles the issues originating from the lack of reliable information about the spatial distribution of the human population in the DRC. Although several population distribution data sets have been delivered at various spatial resolutions and at either continental or global scales, some issues still need to be addressed for a proper modeling of the population distribution in developing countries. This thesis thus proposes an easily interpretable method to derive the spatial distribution of the human population at the DRC nationwide scale (Chapter 2). The method uses the population counts that are available at the coarser scale of the third country administrative level, in order to obtain the human population densities at the spatial resolution of one kilometer. Simulations are conducted to evaluate the potential impact of the method assumptions and results are compared to estimates from available global and continental population data sets. This comparison aims at highlighting the divergences, the inconsistencies, and the further opportunities to improve the population distribution estimations in less developed countries.

The third specific objective is the implementation of a spatial model that can allow forecasting the areas that are more prone to forest cover changes due to their proximity to major towns and big villages. This was achieved in the Chapter 3 by implementing a multimodal accessibility model that uses both the road network and alternative transportation systems to estimate the travel-time from any location of the study area to nearest places of interest. The travel-time is used as an accessibility proxy-variable that is compared with field-based data. The transportation scarcity, the unavailability of reliable information related to the network conditions, and the failure to take into account the use of alternative transportation systems prevented previous approaches to correctly grasp the accessibility issues in DRC. These issues were in turn integrated into the proposed model so as to highlight the potential of a multimodal accessibility modeling for the forest cover change analysis.

The third research question is intimately linked to the previous two since this thesis addresses the following issue:

3. *Can the retrieved spatial variables approximating the main forest cover change drivers for the DRC be jointly analyzed in order to simulate both the quantity and the patterns of past and future forest cover losses at the national scale?*

This question is relevant since several deforestation models that were applied to tropical regions (Mertens and Lambin, 2000; Zhang *et al.*, 2002, 2006; Tollens, 2010; Khoi and Murayama, 2011) did not adequately integrate the proximate causes of forest cover change. This is mostly due to the scarcity of the necessary data to perform a such analysis and the complex relationships between the drivers of change (Nelson and Geoghegan, 2002; Mas *et al.*, 2004). A proper integration of these causes can be achieved by analyzing the interactions between the spatial distribution of the human population and the demand for forest resources and agricultural lands, in conjunction with the accessibility to major towns and big villages. Different scenarios can be considered such as those related to the impact of the population growth on the demand for forest resources.

The last specific objective meets the growing need among both the scientific community and the (inter) national stakeholders for an assessment of the possible futures of DRC forests (Megevand *et al.*, 2013). The Chapter 4 thus simulates the future forest cover changes for a period of thirty years starting from 2005. The developed model uses the information derived from the forest cover change comparison analysis achieved in the Chapter 1, the simulated population figures based on the results obtained in the Chapter 2, and the travel-time estimates — which serve as proxy-information for the proximity of forested lands to major towns and big villages — derived from the model developed in the Chapter 3. The model calibration uses two reference land-cover maps for years 2000 and 2005, while the validation procedure accounted for a third land-cover map for the year 2010. A forest conservation scenario based on the past deforestation trends and a business-as-usual (BAU) scenario that account for population growth are implemented to portray the possible future forest cover losses up to 2035.

Thesis outline

Chapters 1 to 4 constitute the core of this thesis (Figure 1). These are based on scientific papers that have already been published (Chapter 3) and are in finalization for submission (Chapters 1, 2, and 4) in peer-reviewed journals.

The main findings of the thesis, future perspectives for research, and policy implications of findings are summarized and discussed in the conclusion of this document.

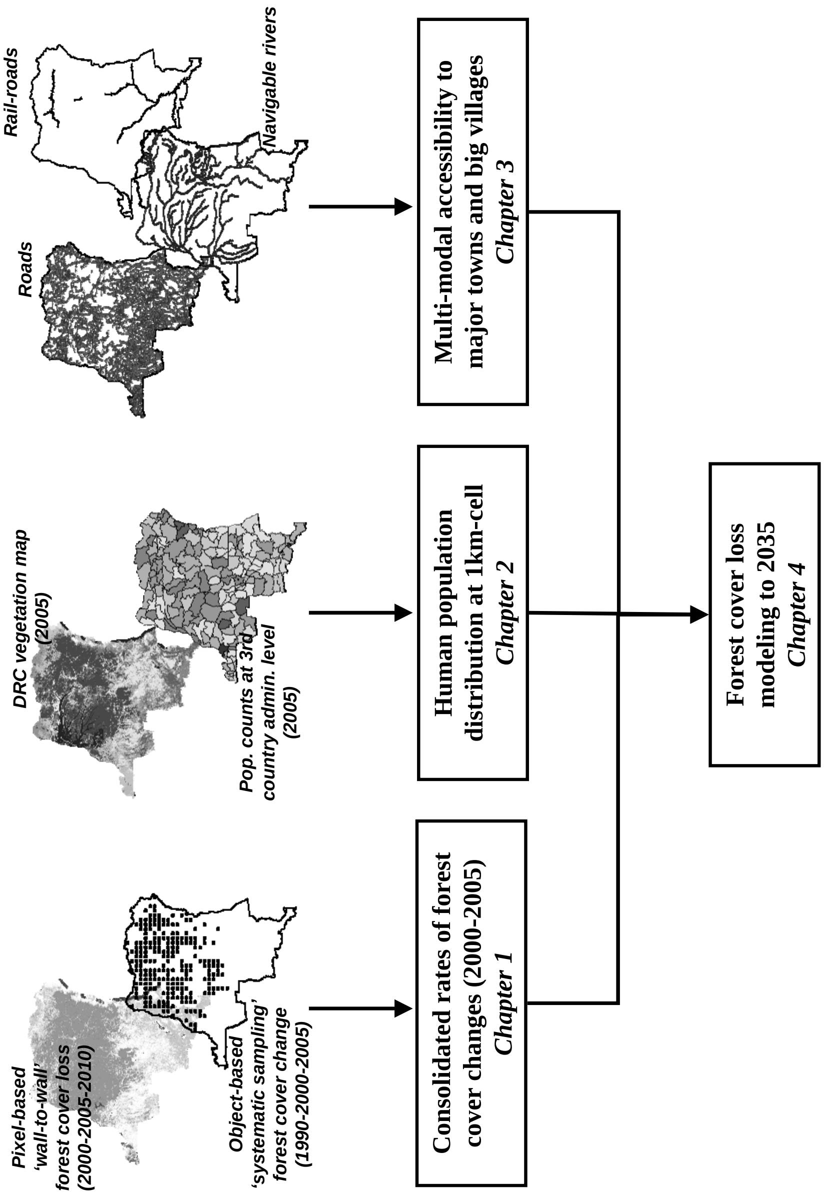


Figure 1: Interactions associated to the main objective of this thesis.

Chapter 1

Combining results from pixel- and object-based digital image processing for forest cover estimates in Democratic Republic of the Congo

Abstract

Monitoring systems have been developed for estimating forest cover and change using remote sensing. These estimations are derived from either exhaustive coverage or sampling designs, using respectively pixel-based and object-based approaches. This chapter explores the sources of discrepancies arising from 'wall-to-wall' pixel-based and systematic sampling object-based approaches in the estimation of the forest cover and change. The proposed method investigates two most recent and Landsat-based products for the Democratic Republic of the Congo (DRC) over the period 2000-2005. Simple legend harmonization rules are used in order to account for the pixel-based information for an object level comparison analysis. The overall agreement found between the two products is 87.2% in surface area for the year 2000 and 86.8% for the year 2005. These high values of agreement are mainly due to the agreement between the two major classes of pixel- and object-based products that are respectively primary forest and tree cover. The implemented comparison analysis is of paramount importance since it contributes to the credibility and the transparency of forest cover change estimates obtained from satellite Earth observation, which is considered as the most efficient approach to monitor DRC forests.

1.1 Introduction

Forest cover change is one of the most dynamic processes of land surface in tropical countries, and satellite Earth observation (EO) has been recognized such as the most effective approach for its monitoring (Achard *et al.*, 2010). Satellite EO thus delivers estimates of forest change at different scales varying from global or sub-continental (Stibig *et al.*, 2007; Duveiller *et al.*, 2008; Hansen *et al.*, 2008a; Bodart *et al.*, 2011), to national and sub-national (Hansen *et al.*, 2008b; Cabral *et al.*, 2011; Eckert *et al.*, 2011; Potapov *et al.*, 2012). The use of satellite imagery requires, however, standards and appropriate methods including atmospheric correction, radiometric calibration, co-registration, compositing, classification, and change detection. At the same time, the sources of remaining uncertainties in image processing are more and more thoroughly documented (McRoberts *et al.*, 2010).

The two main digital image processing methods currently used in tropical forest monitoring are referred as pixel-based (PB) (Hansen *et al.*, 2008a; Lindquist *et al.*, 2008; Potapov *et al.*, 2012) and object-based (OB) (Duveiller *et al.*, 2008; Verhegghen *et al.*, 2010; Ernst *et al.*, 2013) approaches. While the former approach typically transforms the spectral radiance of each single pixel into a land-cover information, the latter usually relies on a preliminary image segmentation in order to delineate sets of adjacent pixels. These are generated by one or more homogeneity criteria, so that output image objects can be characterized by variables that are beyond purely spectral information. Indeed, these variables also include spatial information such as the shape, the distance, the neighborhood, and the topology (Blaschke, 2010; Whiteside *et al.*, 2011).

In addition, the estimates of forest change obtained from any approach can derive either from comprehensive coverage of the entire study area, i.e., exhaustive mapping or so-called 'wall-to-wall' mapping strategy (Hansen *et al.*, 2008a; Haeusler *et al.*, 2010; Broich *et al.*, 2011; Potapov *et al.*, 2012), or from a sampling strategy (Desclée *et al.*, 2006; Duveiller *et al.*, 2008; Hansen *et al.*, 2008b; Beuchle *et al.*, 2011; Rasi *et al.*, 2011; Ernst *et al.*, 2013).

While Congo basin forest cover change estimates obtained from Duveiller *et al.* (2008) were considered as a breakthrough, such change rate estimates are now computed from many sources introducing some confusion for users and decision makers. From a scientific point of view, it is therefore of paramount importance to classify the sources of discrepancies and the various working assumptions behind each approach. Such a study should stem from a theoretical point of view leading to evaluate the methods performance. Recent forest monitoring efforts in central Africa provide a unique opportunity to compare different approaches using the same kind of data over a large area.

The present chapter aims at comparing the two most recent forest cover data sets for the DRC (Potapov *et al.*, 2012; Ernst *et al.*, 2013) over the period 2000-2005. This comparison is hindered since these products were obtained from two different digital image processing methods (PB *vs.* OB) that used different class definitions and two different mapping strategies (exhaustive *vs.* sampling). Both PB and OB approaches make use of labeling procedures that allow enclosing mapping units (i.e., pixels or objects) into categories described such as distinct classes. However, the differences in land-cover typologies makes the inter-comparison of respective outputs quite challenging (Herold *et al.*, 2006; Jung *et al.*, 2006; McCallum *et al.*, 2006; Herold *et al.*, 2008; Pflugmacher *et al.*, 2011). Such a comparison should rely on a-priori translation of the land-cover maps to be compared into a common land-cover language, such as the Land-Cover Classification System (LCCS) (Di Gregorio, 2005). This translation most often requires both local and regional expertise, as well as reference data. As a consequence, the inter-comparison of different land-cover maps should aim at highlighting the divergences and illustrating the possible ways of comparison, since a complete interoperability between different maps is hard — even impossible — to achieve (Herold *et al.*, 2006, 2008).

The approach was to derive from the original pixel-based product an object-based land-cover map (hereafter referred such as the PB-derived map), in order to allow a comparison with the OB product at the object level. This comparison was aimed at providing an assessment of the OB product in regards to the classes depicted in the pixel-based land-cover map.

1.2 Study area and available data

The present study analyzes the two state of the art, full scale, and independent efforts to monitor the forest cover change in the DRC. These are namely the FACET ('Forêts d'Afrique Centrale Evaluées par Télédétection') and the OFAC ('Observatoire des Forêts d'Afrique Centrale') data sets. These were based on the same type of imagery (Landsat images) and shared a common time-frame (2000-2005). They were respectively reported by Potapov *et al.* (2012) and Ernst *et al.* (2013). The analysis is here focused on the DRC dense moist forest extent that was obtained from Verhegghen *et al.* (2012). Indeed, discrepancies between the two products are expected to be more important in dry forests due to higher classification errors and bigger sensitivity to land-cover typology.

1.2.1 Forest mapping and forest cover change estimates

Based on the Landsat compositing methodology developed by Hansen *et al.* (2008a), Lindquist *et al.* (2008) found that using more Landsat acquisitions per path/row by epoch to produce composite mosaics for the DRC resulted in only 4% of pixels remaining unsuitable for change detection analysis. These authors therefore promoted the use of all data present in the Landsat archive — such as achieved by Potapov *et al.* (2012) — to overcome cloud contamination issues, SLC-off gaps, and other prejudicial remote sensing variations. Conversely, the approach developed by Ernst *et al.* (2013) supported that a consistent forest cover change analysis over all Central Africa could be achieved by using *only* pairs (or triplets) of best available Landsat images for each sample location.

The selection of the most suitable image was completed from the Landsat archive as well.

1.2.2 Object-based approach

In the OFAC framework, Ernst *et al.* (2013) applied a semi-automated object-based procedure for deriving annual forest cover change rates by country throughout central Africa. A systematic sampling of pair or triplet multi-year Landsat TM and ETM+ 20 $\times$ 20 km separated by 0.5° interval were selected to obtain estimates of forest change processes over 1990-2000-2005.

The overall object-based approach (Desclée *et al.*, 2006; Duveiller *et al.*, 2008; Verhegghen *et al.*, 2010; Ernst *et al.*, 2013) consisted of four steps that were image pre-processing and pre - interpretation, interpretation by national experts, statistics computation, and accuracy assessment. The pre-processing consisted of mosaicking several images to cover the sample site if needed, co-registering multi-temporal images, calibrating the radiometric data, masking clouds and shadows, and correcting the haze. An automated segmentation procedure (Desclée *et al.*, 2006) was then used to partition the multi-year images into homogeneous objects. These include pixels that are spectrally homogeneous and have a similar trajectory of land-cover change. The segmentation was conducted in two steps. The first step consisted of obtaining small objects (L1) with a minimum size of 1ha, which were used as Elementary Processing Units (Radoux *et al.*, 2008). The pre-interpretation included an unsupervised k-means classification of L1 objects in 20 clusters for each pair or triplet of images. Using the GlobCover product (Defourny *et al.*, 2009) as a reference land-cover map, all L1 objects containing more than 95% of forests in GlobCover were considered as tree cover (TC) training sets. The labeling of remaining L1 objects was based on a Mahalanobis distance computed on Red, NIR, and SWIR spectral bands between TC training sets and each object. L1 objects were then aggregated into bigger objects (L2) to hold only 5% of objects with a size smaller than 5ha for each sample location. The aggregation aimed at reducing the number of polygons to be validated by national experts. Labeling of L2 objects was based on the proportion of L1 labels within each L2 object, resulting in 'pure' or mosaic forest labels. The final land-cover legend was the following: (1) Tree Cover (TC, tree cover proportion is over 70% and tree height should be at least 5 m), (2) High Mosaic (HM, for tree cover proportion comprising between 40 and 70%), (3) Low Mosaic (LM, the tree cover proportion is between 10 and 40%), (4) No Tree Cover (NTC, for tree cover proportion lower than 10%), and (5) Water (WT).

National experts from Congo Basin countries performed the visual interpretation of L2 objects resulting from the automatic classification. These objects were then used to compute forest cover change statistics based on transitions of land-cover classes between 1990-2000 and 2000-2005. Four change processes were characterized and corresponded to: (1) deforestation (transitions from TC, HM, and LM to NTC: TC-NTC, HM-NTC, and LM-NTC), (2) forest degradation (TC-HM, TC-LM, and HM-LM), (3) reforestation (NTC-TC, NTC-HM, and NTC-LM), and (4) forest regeneration (LM-TC, HM-TC, and LM-HM). Given the lack of reference data for the Congo Basin for a strict accuracy assessment, an independent expert conducted a consistency analysis for randomly selected Landsat images, in order to evaluate the accuracy of both land-cover labels and change. An overall accuracy of 85.7% was found for land-cover classes from the 190 polygons reinterpreted.

1.2.3 Pixel-based approach

The FACET processing chain (Potapov *et al.*, 2012) inherited from the methodology developed by Hansen *et al.* (2008a). This was applied to 8,881 Landsat ETM+ images (circa 2000, circa 2005, and circa 2010) with less than 50% cloud cover to deliver a 'wall-to-wall' coverage of the DRC at 60-meter spatial resolution. An analysis of forest cover loss was achieved as well. Two sets of metrics of the periods 2000-2005 and 2005-2010 were also produced.

The overall pixel-based approach comprised four main steps (Potapov *et al.*, 2012) that were the images resampling to 60-m for minimizing false change detection due to sub-pixel misregistration effects and facilitating the further image compositing, the conversion of raw images digital number values to top-of-atmosphere reflectance to minimize differences in ETM+ sensor calibration, the screening of cloud and shadow, and radiometric normalization to reduce reflectance variations between image dates due to atmospheric variations and surface anisotropy. Multi-date image compositing was employed to generate start/end date image mosaics for two six-year time intervals (2000-2005 and 2005-2010). Multi-temporal metrics were generated to evaluate the reflectance variation within the analyzed time interval. A classification tree model was used to map forest cover and types for circa year 2000. For this purpose, an extensive forest/non forest training set was manually created by visual interpretation of the composite imagery representing the earliest cloud-free observation for 2000-2005 time interval. Forest was defined as 30% canopy cover or greater for trees of 5 meters in height or more. To aid the interpretation, freely available QuickBird images and expert information were used as reference materials.

The final output consisted in a thematic map of few classes of both Land Cover and Land Cover Change (LCLCC). The different forest classes included: (1) Primary humid tropical Forests (PF, defined as mature forest with greater than 60% canopy cover) mapped separately within the generic forest mask using a manually created training set, (2) Secondary humid tropical Forest (SF, defined as regrowing immature forest with greater than 60% canopy cover) mapped from a supervised classification model using a manually created training set in addition to multiple criteria such as geographic distribution, proximity to remain forest patches, tree composition and canopy density, and (3) Woodlands (WO, defined as forest cover with greater than 30% and less or equal to 60% canopy cover) corresponding to remaining forest cover after mapping primary and secondary humid tropical forest classes. Permanent water bodies (WT) were mapped separately. Gross forest cover loss from 2000 to

2005 was mapped within the 2000 forest mask, and forest cover loss between 2005-2010 was mapped within the remaining forest of year 2005. Forest cover loss mapping included creating two set of manually interpreted training sites using visual analysis of forest cover change for the 2000-2005 and 2005-2010 intervals. Two separate classification models were built upon these training data and the images composites of year 2000, 2005, and 2010 were used for visual interpretation of change training data. The 2005 composite was used as start point for the 2000-2005 interval, and the endpoint for the 2005-2010 interval. Forest cover loss mapping did not make any distinction between the different change types such as agriculture clearings even followed by forest regrowth within the same time interval, logging, fire, etc.

Key properties from OB and PB products are summarized in the Table 1.1. Pixel-based approaches are most common satellite imagery processing methods which consist of transforming spectral radiances in land-cover information at the pixel level. However, since Blaschke and Strobl (2001) raised the provocative question *what's wrong with pixels?*, and even shortly before (Cracknell, 1998), there is a growing interest towards 'beyond pixels' applications (Blaschke, 2010). These rely usually on preliminary image segmentation to provide building blocks of pixels which are generated by one or more homogeneity criteria. Output image objects are therefore characterized by variables that are beyond purely spectral information and include spatial information such as shape, distance, neighborhood, and topology (Blaschke, 2010; Whiteside *et al.*, 2011).

Table 1.1: Key figures from OFAC and FACET data sets.

	OFAC	**FACET**
1. Study area	Central Africa	DRC
2. Mapping strategy	Systematic sampling design	'Wall-to-wall' mapping
3. Initial data set	708 Landsat scenes	8,881 Landsat scenes
3. Time-frame	1990 - 2000 - 2005	2000 - 2005 - 2010
4. Image processing	Object-based (OB)	Pixel-based (PB)
5. MMU	5 ha-object	0.36 ha (60 x 60m)-pixel
6. Land cover classes	- Tree cover (TC) - High Mosaic (HM) - Low Mosaic (LM) - No Tree Cover (NTC) - Water (WT)	- Primary Forest (PF) - Secondary Forest (SF) - Woodlands (WO) - Non Forest (NF) - Water (WT)
7. Change detection	From object transitions between two time-periods	From multi-temporal spectral metrics
8. Estimated annual rates (%) for 2000-2005	- Deforestation: 0.32±0.05 - Degradation: 0.16±0.03 - Reforestation: 0.10±0.03 - Regeneration: 0.04±0.02	Forest cover loss: 0.22
9. Final product validation	Carried out by national experts	NA

Pixels are Elementary Measurement Units (EMU) common to both pixel-based and object-based approaches. Through image analysis processing (resampling, mosaicing, segmentation, etc.), an Elementary Processing Unit (EPU) is obtained that could theoretically correspond to a single pixel, while is most of often made of several pixels in the object-based approach. In addition, the minimum size for objects labeling is the Smallest Legend Unit (SLU) which corresponds to the thematic constraint for meeting all the classification process requisites (Radoux *et al.*, 2008). This latter still different than the Minimum Mapping Unit (MMU) which is set according to cartographic standards and corresponds to the smallest element for mapping representation. Therefore, PB and OB products are most often not straightforwardly comparable since their underlying methods are conceptually different. Indeed, single pixels delivering the PB product are upscaled at the object level in the OB approach. The two approaches are thus based on different units (i.e., EPU, SLU, and MMU), that modified their respective analysis scale. Observed dynamics at the elementary unit of analysis (i.e., pixel or object) are consequently different which leads to dissimilar outputs with both different legends and errors that are specific to each method.

The PB product was obtain from an exhaustive ('wall-to-wall') mapping of forest cover and change, while the OB product was based on a systematic sampling. To compare deforestation rates obtained from both mapping strategies, Tucker and Townshend (2000) implemented an experiment design to estimate the number of Landsat samples needed to achieve a given percentage (i.e., ±10%, ±20%, and ±50%) of actual deforestation 90% of the time for Columbia, Peru, and Bolivia. They found that it was necessary to sample 70-90% of the total area (i.e., almost the entire area) to obtain a ± 20% estimate of deforestation 90% of the time. These authors concluded that random sampling of Landsat imagery was unlikely to yield accurate estimates of deforestation, particularly for areas such as humid tropical forests where deforestation is spatially concentrated. However, Czaplewski (2003) demonstrated that conclusions from Tucker and Townshend (2000) were no longer valid out of the range of their experimental design. In fact, Czaplewski (2003) found that at equal sampling fraction, i.e., 10%, sample-based estimate of deforestation was likely to be near its true value by increasing the extent of the geographic area being analyzed, and the resulting sample size, for the same type of spatially concentrated deforestation. This author concluded that a sample of Landsat scenes can provide reliable estimates of deforestation if sample size is sufficiently large and stratified sampling could yield equally precise results with smaller sample sizes. In addition, Duveiller *et al.* (2008) suggested that the pan-tropical survey of forest cover changes realized by the FRA (FAO, 2001) based on a stratified random sampling scheme using 117 sampling units whose size corresponds to a complete Landsat scene, i.e., 185×185 km, could be improved significantly by using small image extracts as sampling units, and having them systematically (rather than randomly) distributed over the forest domain. These authors realized a survey of 10×10 km Landsat samples systematically-distributed every 0.5° over the whole forest domain of Central Africa, corresponding to a sampling rate of 3.3%, to estimate deforestation at regional, national and landscape levels. To test the impact of the sampling rate on the stability of the results, Duveiller *et al.* (2008) lowered the sampling rate to 0.8% by dividing the entire population into four subsets, each with one sample site every square degree. These authors suggested using more intensive sampling strategy since they found that sampling rate played a critical role for getting representative and robust estimates. Indeed, Congo Basin annual deforestation estimates varied considerably from 0.15% to 0.27% for a sampling rate of 0.8%, with very broad confidence interval, while the corresponding estimate at sampling rate of 3.3% was 2.45%±0.55%. Beyond the controversial debate regarding the best option for accurately estimating tropical deforestation, i.e., 'wall-to-wall' mapping *vs.* sampling of Landsat scenes, it can however be stated that the former approach has the advantage of providing data on the spatial distribution of forests for geographical analysis while stratified sampling has a greater contribution for *hot spots* deforestation monitoring.

The PB data set mapped the gross forest cover loss from which it is impossible — without analysis of further land-use — to predict if clearing represents deforestation or temporary tree cover removal since no distinction was made between change types in the analysis. The two levels of image segmentation such as conducted in the OB approach allowed mapping four different forest cover change processes, due to the possibility to detect variant change dynamics in L2 objects. Distinguishing deforestation from forest degradation is of paramount importance since this latter is the more extensive and often more insidious process (Putz, 2010). Measuring forest degradation from satellite EO is, however,less straightforward in comparison to deforestation. This is partly due to the forest degradation definition itself stating that it corresponds to the reduction of the capacity of the forest to provide goods and services (FAO, 2010). The forest degradation such as measured by the OB product is based solely on detecting change in forest area and do not take into account the qualitative forest degradation, i.e., change in species composition, forest disease, forest disturbances, etc. In addition, none of the PB and OB products were validated using field data that lacks for DRC, making the accuracy assessment a common weakness to both approaches. Involving national experts from Congo Basin countries in the photo-interpretation of the OB product certainly contributed to both capacity building and national ownership of the products, although it remains doubtful that all experts were equally trained and experienced. This is particularly true since Ernst *et al.* (2013) did not provide any information about the impact of possible experts subjective choices on the interpretation quality.

1.3 Data sets comparison

From the 547 sampling sites processed by Ernst *et al.* (2013), only 233 samples distributed over the DRC moist dense forest extent were used in the current analysis (See Figure 1.2). They thus correspond to a sampling rate of 7.5% for a total extent of about 1,232 thousands km^2. Some missing samples were found and corresponded to both core unchanged forest areas (i.e., the confluence of Equateur, Kasaï-Occidental, and Bandundu Provinces corresponding to the location of the Salonga National Park) and forest change dynamic places such as artisanal mining in southern Province Orientale or clearings in and near the Virunga National Park. These missing samples were due to data unavailability for the period 2000-2005, while they were available for the period 1990-2000. Since these sampling sites more likely remained with similar forest (un)change dynamics between 1990-2005, a way to assess the effect of these gaps on change estimation and confidence interval is to compute the 1990-2000 and 1990-2005 forest cover change statistics with these gaps and compare the results to the original full sampling coverage for the same periods.

For the purpose of the comparison analysis, the 2000 and the 2005 PB land-cover maps were derived from the original FACET data set. The comparison analysis was performed at the object level. Each of the OB polygon (i.e., object) found in the 233 samples was processed in order to obtain PB-derived objects. These correspond to OB delineated objects labeled using the OB classification scheme, based on the distribution of PB pixels withing each object. This was achieved thanks to simple legend harmonization rules between the two classifications (Figure 1.1 - *Step I* and Section 1.3.1). The legend harmonization therefore allowed providing the label corresponding to each object based on the PB information (i.e., PB-derived objects). PB-derived and original objects were afterwards compared per year (2000 and 2005) to detect disagreements (Figure 1.1 - *Step II* and Section 1.3.2). The third step of the comparison analysis (Figure 1.1 - *Step III* and Section 1.3.3) aimed at selecting a representative sample of disagreed objects for visual analysis (Figure 1.1 - *Step IV* and Section 1.3.4) to retain only the most plausible label, i.e., the consolidated label, per year for each object (Figure 1.1 - *Step V* and Section 1.3.5). The computation of forest cover change rates with their confidence intervals, which refer to the sampling variability, concluded the comparison analysis. The overall comparison analysis is structured such as presented in the Figure 1.1.

1.3.1 Legend harmonization

The legend harmonization rules (Figure 1.1 - *Step I*) were based on the ratio ($Prop$) between the total area of PB classes lying within objects' boundaries and the object area (see Expression 1.1). Thus, PB-derived objects were classified such as 'Tree Cover' for $Prop \geq 0.7$ of forest classes. 'PB-derived High/Low mosaics' corresponded to $Prop$ values comprised respectively between 0.7-0.4 or 0.4-0.1. For $Prop \leq 0.1$, PB-derived objects were classified such as 'PB-derived no tree cover'. These thresholds values correspond to those used for the classes' definition in the original OB data set.

$$Prop = \frac{PF_{area} + SF_{area} + (0.45 \times WO_{area})}{Objectarea} \tag{1.1}$$

where PF_{area}, SF_{area}, and WO_{area} areas stand respectively for Primary

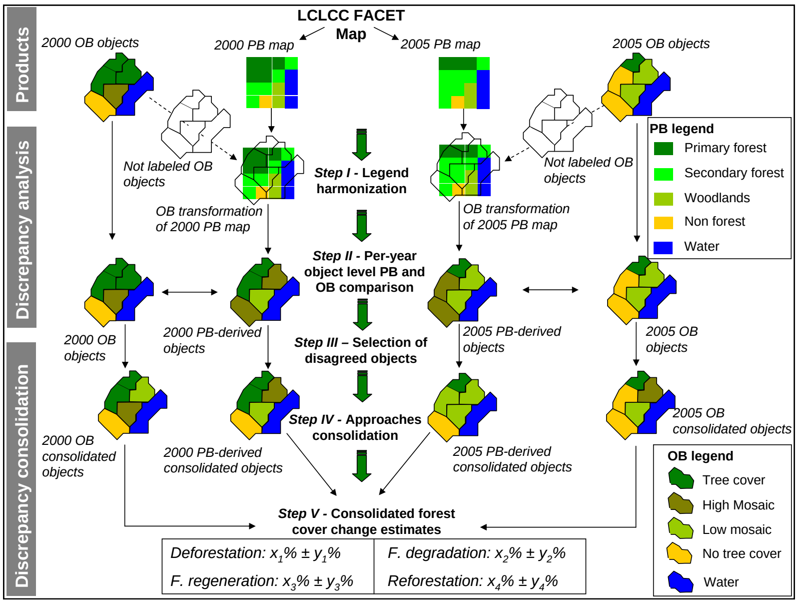


Figure 1.1: Comparison methodology of PB-derived and OB data sets at the object level.

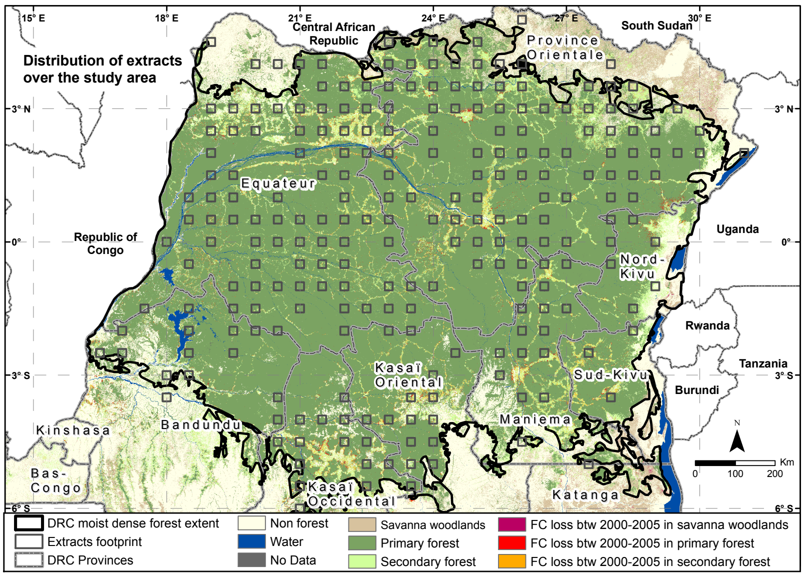


Figure 1.2: Distribution of the 233 extract units overlaid on the FACET forest cover/loss map.

Forest, Secondary Forest, and Woodland areas that were directly retrieved from the 2000 and 2005 PB land-cover maps. The correction factor 0.45 in the Expression 1.1 refers to the central value of canopy cover proportion in the woodlands' definition (30-60%).

The total forest cover area (FC) for both PB-derived and OB data sets were computed according to the following Expression 1.2:

$$FC = \frac{TC_{area} + (0.55 \times HM_{area}) + (0.25 \times LM_{area})}{0.075} \tag{1.2}$$

where TC_{area}, HM_{area}, and LM_{area} correspond to classes areas. Correction factors 0.55 and 0.25 in the Expression 1.2 are respectively central values of forest cover proportions in definitions of high mosaic (40-70%) and low mosaic (10-40%) (Ernst *et al.*, 2013). The denominator 0.075 corresponds to the sampling rate (7.5%).

1.3.2 Comparison of PB-derived and OB data sets

PB-derived and OB data sets were compared for years 2000 and 2005 for the whole 233 extracts (Figure 1.1 - *Step II*). This was achieved by cross-tabulating the two products to highlight agreements and disagreements between the two data sets. In the resulting contingency tables, agreement cases were found in the diagonal cells, while the different cases of disagreement corresponded to non diagonal elements.

To further evaluate the different disagreement cases, a random selection (see Section 1.3.3) was performed since validating the large number of disagreed objects would be time consuming. This assessment, aiming at obtaining the true label for each disagreed object, was achieved through visual interpretation by overlapping selected disagreed objects on the corresponding Landsat images for each year.

1.3.3 Selection of disagreed objects

Given the various different size of disagreed objects to be assessed, it is of major concern to give equal chance to both small, medium, and bigger disagreed objects to be selected. Radoux *et al.* (2011) proposed a framework to achieve such an unbiased selection, while these authors also found that for a reduced number of objects (i.e., N < 500), it was very difficult to obtain a representative subset of the different objects' size. In this case, the entire set of disagreed objects was analyzed in the current analysis.

For N > 500, the number of selected objects (n) for each disagreement case corresponded to the positive solution of the below quadratic Equation (see Expression 1.3), which was retrieved from Radoux *et al.* (2011):

$$\left(\frac{c_v^2 N}{N-1}\right) n^2 + \left(\frac{CI^2 N}{p_0\,(1-p_0)\, z_{1-\alpha/2}^2} - \frac{N^2 c_v^2}{N-1}\right) n + N = 0 \tag{1.3}$$

where N is the total number of objects in a given disagreement case and c_v is the corresponding coefficient of variation. $c_v^2 = \sigma^2/\mu^2$, with σ being the N objects areas standard deviation, and μ being the corresponding mean. CI corresponds to the half of the significance level (i.e., $CI = 0.025$), and $z^2_{1-\alpha/2}$ corresponds to the z-value for a given confidence level.

For $N > 500$ and two-sided confidence level of 95% ($\alpha = 0.05$), $z_{\alpha/2}$ equals to 1.96. p_0 in turn corresponds to the probability of an object to be correctly classified according to the OB data set. While this information was not provided by Ernst *et al.* (2013), a value of $p_0 = 0.5$ was retained for the current study. This value can be considered such as conservative since it corresponds to the maximum value of n according to the Expression 1.3.

For disagreement cases showing a large c_v (typically greater than 1.5), Radoux *et al.* (2011) found that the efficiency of the object-based sampling was reduced (compared to point-based sampling) because of the low probability to sample one of the large objects. In these cases, a point-based sampling was therefore used. The points were randomly generated within the extent of objects affected by the targeted disagreement. For $p_0 = 0.5$, the number of random points was 1,537 (Radoux *et al.*, 2011). For the further visual analysis, a 1-kilometer buffer around each point was considered to allocate the land-cover class corresponding to the point.

1.3.4 Visual interpretation of disagreed objects

The selected disagreed objects were overlaid on Landsat false color composite for the visual interpretation (Figure 1.1 - *Step IV*) to obtain the so-called consolidated label. This can correspond to the label already provided by one of the two products, i.e., OB or PB-derived, or a different label in case both products mislabeled the actual object. The images used were those that served for the OB forest cover estimates. The visual interpretation evaluated about 18,600 sampling units from 200 different Landsat extracts (Appendix A), half for year 2000 and half for year 2005.

1.3.5 Forest cover change rates consolidation

The consolidated label resulting from the visual interpretation was used as reference information to perform the accuracy assessment (AA) of both OB (AA_{OB}) and PB-derived ($AA_{PB-derived}$) products. The assessment accuracy for analyzed polygons was computed according to the following Expression (Radoux *et al.*, 2011) :

$$AA = \frac{\sum_{i=1}^{n} C_i S_i + \sum_{i=1}^{n} C_i/n \sum_{i=i+1}^{N} S_i}{\sum_{i=1}^{N} S_i} \tag{1.4}$$

where possible values for C are 1 for objects correctly classified according to the visual analysis and 0 for objects wrongly classified. S represents the object area.

The assessment accuracy for point-based visual analysis was computed as follows (Radoux *et al.*, 2011):

$$AA = \frac{\sum_{i=1}^{n} C_i}{n} \tag{1.5}$$

where n is the number of points within the corresponding sample.

In order to select the best product to be used per disagreement case, $AA_{PB-derived}$ and AA_{OB} were compared using a formal one-tailed Student's t-test for paired observations at $\alpha = 0.05$.

The null hypothesis was formulated as:

$H_0 : \mu_{AA_{PB-derived}} = \mu_{AA_{OB}}$

while the alternative hypothesis was:

$H_0 : \mu_{AA_{PB-derived}} \neq \mu_{AA_{OB}}$

The statistic (t_{calc}) used was computed according to the Expression:

$$t_{calc} = \frac{\overline{D_{AA}}}{\sqrt{\frac{S^2_{D_{AA}}}{n}}} \tag{1.6}$$

with $\overline{D_{AA}}$ being the mean of differences between $AA_{PB-derived}$ and AA_{OB} and n being the number of objects selected for each disagreement case. The variance of differences ($S^2_{D_{AA}}$) was computed as follows:

$$S^2_{D_{AA}} = \frac{1}{n-1} \sum_{i=1}^{n} (D_{AA} - \overline{D_{AA}})^2 \tag{1.7}$$

p-values resulting from the t_{calc} statistic allowed to construct rules to consolidate PB-derived and OB maps (Figure 1.1 - *Step IV*). For *p-values* < 0.05, the null hypothesis was rejected and disagreed objects were re-labeled using the consolidated label. If the null hypothesis could not be rejected, the corresponding disagreement remained. After applying consolidation rules re-labeled objects were used to compute annual rates of forest cover change between 2000-2005 (Figure 1.1 - *Step V*). The computation was based on the object transitions between the two years.

For the change rates computation, object areas from each 20 × 20 km extracts were normalized to take into account the latitude effect (because of the geographic projection) and the exact date of image acquisition. To reflect the forest cover loss, weights were given to each process according the actual forest cover proportion corresponding to each class (Ernst *et al.*, 2013). For example, the total area converted from LM (with *Prop* between 10% and 40%) to NTC is multiplied by 0.25 since the mean tree cover proportion of LM objects is 25%. Statistics obtained from each 20 × 20 km extracts were summarized at the DRC moist dense forest extent. Their confidence intervals were computed to inform about the sampling uncertainty.

1.4 Results

1.4.1 Area-based comparison of PB-derived and OB data sets

Results from the maps comparison after legend harmonization for the year 2000 are presented in the Table 1.2. The overall agreement between the two maps was high and reached 87.2%. A similar trend was observed for the year 2005 which maps' comparison exhibited an overall agreement of 86.8%. The results were therefore consistent between the two years. These high values are mainly due to the agreement between the major classes of the pixel- and object-based approaches that are respectively primary forest and tree cover classes.

The forest cover areas within the 20 $\times$ 20 km extracts for the year 2000 obtained from both PB-derived and OB products were respectively 82,720 km^2 and 75,900 km^2. These corresponded respectively to a total forest cover area of 1,102,930 km^2 and 1,012,000 km^2 in the DRC humid forest extent. Corresponding estimations for the year 2005 were 81,620 km^2 and 74,890 km^2 in the 233 samples, which were equivalent to 1,088,270 km^2 and 998,530 km^2 of total forest cover area.

Table 1.2: Area-based contingency matrix for year 2000 (values in km^2). The $*$ indicates the total forest cover area (computed from the numerator of the Expression 1.2) within the 20×20 km blocks analyzed.

	OB classes							
PB-derived	TC	HM	LM	NTC	WT	Total	**Forest area**	**Agreement**
TC	73,013.3	1,720.1	1,498.7	5,006.5	102.5	81,341.1		90%
HM	813.3	138.2	136.9	1,157.3	25.6	2,271.3	82,720$*$	6%
LM	80.2	10.7	20.2	388.8	0.5	500.4		4%
NTC	320.4	205	252.3	6,518.1	6.4	7,302.2	–	89%
WT	57	2.9	2.3	43.9	919.3	1,025.4	–	90%
Total	74,284.2	2,076.9	1,910.4	13,114.6	1,054.3	92,440.4	–	–
Forest area$*$	75,900$*$			–	–	–	–	–
Agreement	98%	7%	1%	50%	87%	–	–	87.2%

Highest agreement values were observed for the tree cover and the water classes for both PB-derived and OB classes. High and low mosaics agreements for the two products were, however, very low. Indeed, 'OB high and low mosaic' classes were made of 40% of the two PB primary and secondary forests (Figure 1.3). 'PB-derived high/low mosaics' were in turn made of around 70% of PB woodlands for both years (Figure 1.4). OB and PB-derived high and low mosaics classes therefore exhibited a poor matching by comparison such as depicted in the Table 1.2.

The agreement of the 'OB no tree cover' class was 50% in 2000 and 51% in 2005, while the corresponding PB-derived agreement for the same class was estimated to be 89% for both years (Table 1.2). Indeed, from Figure 1.3, it can be observed that about 40% of 'OB no tree cover' areas corresponded to PB secondary forest and woodlands. These areas were labeled either as 'PB-derived tree cover' or 'high mosaic' after the legend harmonization, which led to the lower agreement observed for the 'OB no tree cover' class (Table 1.2).

Two types of disagreements represented up to 40% of the total disagreements for the two time-periods. These were the 'PB-derived no tree cover'—'OB tree cover' and the 'PB-derived tree cover'—'OB-no tree cover'. They were mainly due to the fact that most of the rural complex areas (= 'OB no tree cover') were appear to be mapped by the FACET data set as secondary forest (Figure 1.6a).

To explore further the possible sources of the diverse disagreements found, PB-derived objects containing 90% or more of one of the PB forest classes were compared to the corresponding OB classes (Figure 1.5). These specific PB-derived objects were respectively denoted such as 'PB-derived pure primary forest', 'PB-derived pure secondary forest', and 'PB-derived pure woodlands'. The pure primary forest objects corresponded very well with the 'OB tree cover class'. The cross-comparison for the secondary forests and the woodlands were in turn more complex. 'PB-derived pure secondary forest' included around 60% of 'OB no tree cover', 25% of 'OB-high and low mosaics', and 15% of 'OB tree cover' for both 2000 and 2005. The 'PB-derived pure woodlands' were made of 50% of 'OB no tree cover', 45% of 'OB tree cover', and the remaining 5% corresponded to 'OB high and low mosaics'.

1.4.2 Analysis of disagreed objects

The discrepancies observed in the Table 1.2 and in Figures 1.3 and 1.4 were analyzed through an extensive visual interpretation exercise. The number of sampling units analyzed for each disagreement case is presented in the Table 1.3, along with the corresponding c_v values.

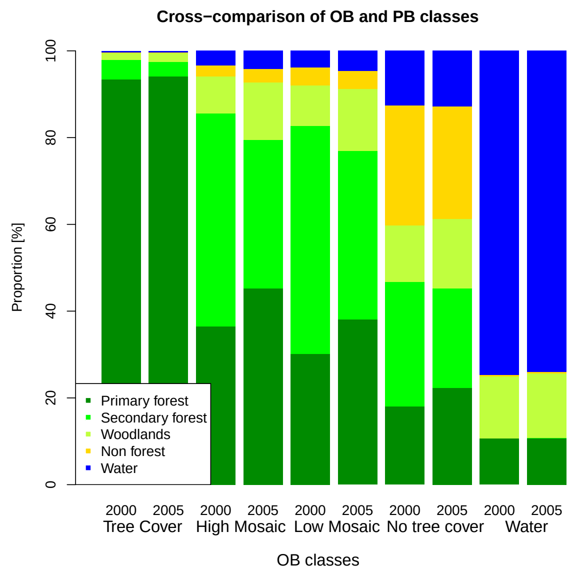


Figure 1.3: Proportion of PB classes within OB classes

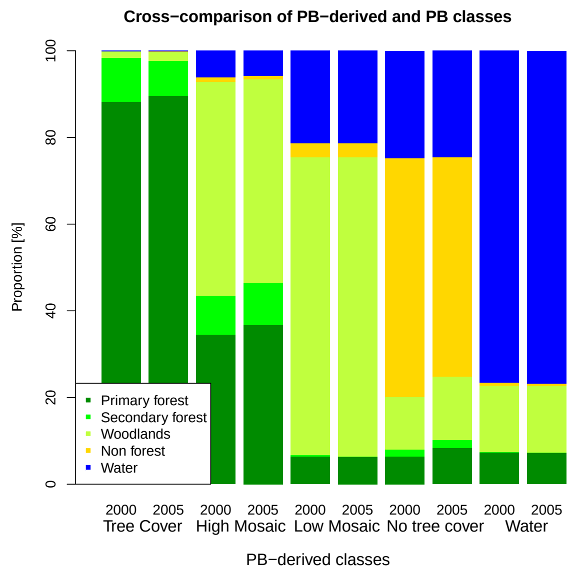


Figure 1.4: Proportion of PB classes within PB-derived classes.

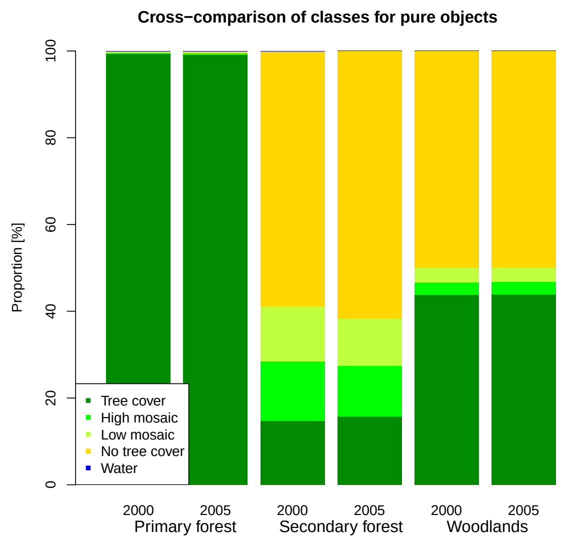


Figure 1.5: Comparison of 'pure' objects with OB classes. 'Pure' objects are those made at 90% or more of their area by a single PB class.

Table 1.3: Number of objects required for the disagreement analysis by visual interpretation. Coefficients of variation are provided between brackets. The '*' mark indicates disagreements with less than 500 objects.

	OB classes				
PB-derived	TC	HM	LM	NTC	WT
Tree Cover (TC)	-	$1{,}752_{(1.46)}$	$1{,}600_{(1.25)}$	$1{,}537_{(1.55)}$	$1{,}537_{(1.63)}$
High Mosaic (HM)	$1{,}537_{(1.88)}$	-	$752_{1.17}$	$1.537_{(1.62)}$	$397^{*}_{(1.36)}$
Low Mosaic (LM)	$756_{0.95}$	$82^{*}_{(1.24)}$	-	$1{,}362_{(1.26)}$	$6_{(0.60)}$
No Tree Cover (NTC)	$1{,}523_{1.15}$	$977_{(1.07)}$	$1{,}013_{(1.02)}$	-	$174^{*}_{(1.96)}$
Water (WT)	$1{,}100_{(1.63)}$	$80^{*}_{(0.81)}$	$58^{*}_{(0.76)}$	$857_{(1.18)}$	-

The Figure 1.6 provides some illustrations of the disagreements visually analyzed. These included the 'OB-no tree cover'—'PB-derived tree cover' that most often corresponded to rural complex areas, which were mapped such as 'PB secondary forest' (Figure 1.6a). Disagreements of type 'PB-derived high or low mosaic'—'OB-tree cover' were found for some of the dry forest patches mapped such as 'PB woodlands' (Figure 1.6b). The Figure 1.6c represents some of the areas that were wrongly classified such as 'OB tree cover', while they clearly lied within the non forest classes from both the Landsat image and the FACET map. This can be attributed to the interactive labeling error in the OFAC data set (Ernst *et al.*, 2013). These objects were labeled such as 'PB-derived no tree cover' following the legend harmonization rules used.

1.4.3 Consolidated estimates of forest cover change

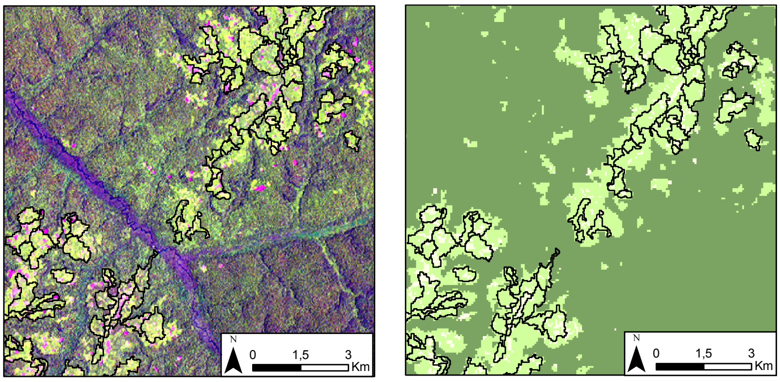
The formal t-test that compared PB and OB assessment accuracies per disagreement case obtained from the visual screening, allowed consolidating 87% of disagreements areas for year 2000 and 86% for year 2005 at the 95% confidence level. Resulting *p-values* are presented in Appendix A, along with the number of different Landsat images analyzed for each disagreement case.

'PB-derived NTC' — 'OB TC' disagreements were corrected for the benefit of the PB approach, while 'PB-derived TC' — 'OB NTC' disagreements were corrected using the OB approach.

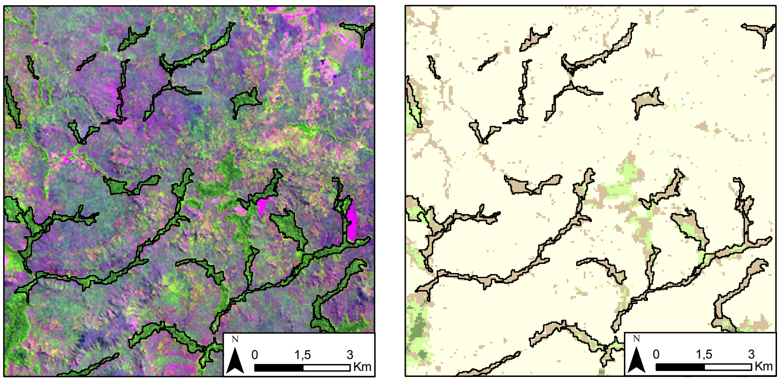
'PB-derived HM' — 'OB NTC' disagreements were consolidated using the OB class (i.e. NTC). This disagreement case referred mainly to objects made essentially of the 'PB secondary forest' (and/or woodlands) class, which were labeled as 'PB-derived HM' after legend harmonization.

The visual screening did not allow consolidating certain disagreements (namely HM-TC, LM-TC, WT-TC, LM-HM, NTC-HM, WT-HM, LM-NTC, LM-WT, and NTC-WT) which assessment accuracies exhibited *p-values* > 0.05, preventing the rejection of the null hypothesis at 95% of confidence level.

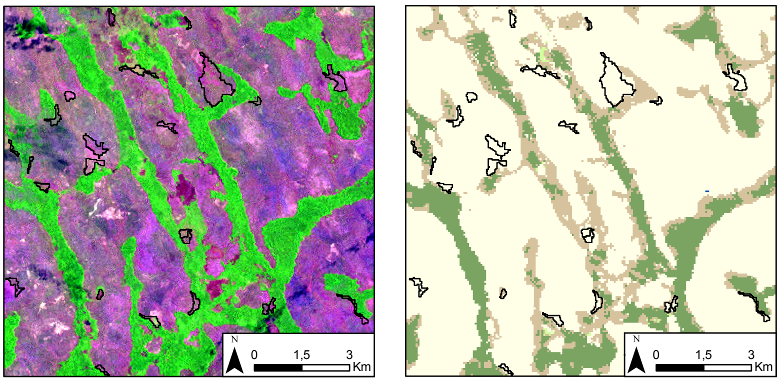
The consolidated rates of forest cover change obtained are presented in the Table 1.4. Since the PB data set did not map the vegetation regrowth, it was not possible to compute the reforestation nor the forest cover regeneration from PB-derived objects.



(a) Landsat image 0°N;24.5°E of date 03/03/2001. 'OB NTC' corresponding to rural complex patterns classified as 'PB SF'.



(b) Landsat image 4.5°N;19°E of date 05/03/2001. 'OB tree cover' classified as 'PB woodlands'.



(c) Landsat image 4°N;20.5°E of date 06/04/2000. 'OB' wrongly classified as tree cover and corresponding to 'PB non forest'.

Figure 1.6: Disagreements between OB and PB-derived products. Delineated objects are overlaid either on the Landsat image (RGB=SWIR-NIR-Red) and on the derived FACET map for year 2000 (PF in green, SF in light green, WO in brown, and NF in white).

Table 1.4: Consolidated annual rates of gross forest cover change.

	Deforestation	F. degradation	Reforestation	F. regeneration
PB-derived	0.18% ± 0.04%	0.09% ± 0.04%	NA	NA
Ernst *et al.* (2013)	0.32% ± 0.05%	0.16% ± 0.03%	0.10% ± 0.03%	0.04 ± 0.02 %
Consolidated estimates	**0.29% ± 0.05%**	**0.16% ± 0.03%**	**0.08% ± 0.02%**	**0.04% ± 0.01%**

1.5 Discussion

The analysis carried out in this chapter aimed at comparing the latest forest cover data sets for the DRC. These were based on different mapping strategies ('wall-to-wall' *vs.* systematic sampling) and used different forest classes, as well as different processing units (pixels *vs.* objects). The two approaches shared a common time-frame (2000-2005) and processed Landsat images to monitor the forest cover change in the DRC.

In order to achieve the inter-comparison of these two products, the working assumption was to convert the PB data set into an OB land-cover map, using the objects delineated by the object-based approach. The proportion of the PB classes lying within each object was used to obtain the PB-derived land-cover map, based on the thresholds values used for the OB classification scheme. An object-level comparison was therefore conducted that allowed assessing the original OB map against the information provided by the PB 'wall-to-wall' mapping strategy.

The high overall agreement between the PB-derived and the OB products was mainly due to the consistency between the OB and the PB-derived tree cover class, which represented up to 85% of the all set of classes. This consistency has to be attributed in a large extent to the proportion (about 90%) of the 'PB primary forest' class lying within both OB and PB-derived objects for the two years of analysis. There is therefore an almost perfect equivalence between the 'OB tree cover' and the 'PB primary forest class', at least for the 233 Landsat samples analyzed.

In turn, 'PB-derived high (and low) mosaics' classes exhibited a poor agreement with the corresponding 'OB high (low) mosaics'. Though, the total area of these classes represented only a small proportion (less than 10% for both years) of the all classes' area. This discrepancy can be attributed to the objects of the 'OB no tree cover' class corresponding to rural complex areas (Figure 1.6a). These were considered such as non forest in the OB data set, while they were mapped within the 'PB secondary forest' class. This is confirmed by the proportion (about 25%) of the whole 'OB no tree cover' class corresponding to the 'PB' secondary forest for both 2000 and 2005 (see Figures 1.3). An option to reduce this type of disagreement could be to merge the two 'OB high/low mosaic' classes into one mosaic class. A single mosaic class could compare well with the information retrieved from the PB product.

According to Vancutsem *et al.* (2008) and Verhegghen *et al.* (2012), the rural complex class refers to fallows, vegetation regrowth, food crops areas, agriculture areas in forest, and village plantations. These land-cover dynamics mostly correspond to shifting cultivation, which is one of the major causes of the forest cover change in the DRC (Tollens, 2010; Defourny *et al.*, 2011). The percent of tree cover in the objects lying within this rural complex class can in fact be lower than 10% of the minimum mapping unit of 5-ha. The OB approach will thus classify this type of areas such as 'OB no tree cover'.

In addition, shifting cultivation mostly corresponds to isolated pixels ('pepper - and - salt' pattern) belonging to the PB forest cover loss within the secondary forest class (Mayaux *et al.*, 2005). These types of forest clearings can theoretically overestimates the gross deforestation if no distinction is made between permanent forest conversion (deforestation) and temporary conversion by shifting cultivators (Angelsen, 1999). This distinction can be made feasible by up-scaling these pixels at the object level — such as achieved in the present chapter — in order to obtain a gradient of change for each object. According to the analysts' thresholds, this gradient can be cut off into different phases of forest cover changes. These can range from forest degradation to forest clearing, or from forest regeneration to forest regrowth such as illustrated in DeFries *et al.* (2007). Shifting cultivation is in turn difficult to handle using the PB approach since the FACET data set did not map the vegetation regrowth.

Besides, the disagreements between the 'PB-derived no tree cover' and the 'OB tree cover' were made of objects with areas close to the minimum mapping unit (5.15 ha $\pm$ 0.15 ha) of the OB approach. These objects were at 90% or greater made by the 'PB non forest' class. This type of disagreement clearly indicates the possible improvement of the OB product using the information from the 'wall-to-wall' forest cover mapping, since single changed pixels from forest to non forest are expected to be well detected using the PB approach.

In turn, disagreements of types 'PB-derived tree cover' — 'OB no tree cover' and 'PB-derived high mosaic' — 'OB no tree cover' were made of bigger objects (respectively 11.59 ha $\pm$ 0.17 ha and 9.59 ha $\pm$ 0.36 ha). These correspond mainly to rural complex areas, while 40% - 70% of their area is mapped as 'PB secondary forest'.

Other disagreements were found for some patches of dry forests that were pooled into the 'PB woodlands' class, while they were considered such as 'OB tree cover' (Figure 1.6b). This could just be an artifact in the FACET data since similar linearly-shaped features (Figure 1.6c) were classified such as 'PB primary forest' by the same data source. In addition, some of the disagreements of type 'PB no tree cover' — 'OB tree cover' (see Figure 1.6c) were likely to be interactive labeling errors rather than wrong classifier rules in the semi-automated OB processing chain.

Overall, one of the major benefits to transform the PB data sets into objects is the straightforward detection of the objects that have been wrongly classified in the OB product. These can be used for the accuracy assessment of the OB approach. Indeed, objects that delineate homogeneous land-cover patterns can be overlaid on colored compositions of satellite images' time series. Analysts can therefore meaningfully distinguish and quantify the changes occurring within objects that were undisturbed at previous time periods.

It is afterward possible to spatially dis-aggregate the rate of forest cover change (see Figure 1.7), since the change dynamics can be of variable intensity within the object boundaries. This direct mapping of forest cover change rates is more readable and informative than a PB LCLCC map, which only highlights changed pixels. In addition, 'salt-and-pepper' structure corresponding to shifting cultivation is thus handled more efficiently at the object level since the various small clearings lying within object boundaries are quantified, which allow the interactive labeling to discriminate various forest cover change processes.

The inter-comparison of maps also highlighted some of the inconsistencies found in the FACET data, which should be systematically validated using field-based data or very high resolution images. This is the case for the vegetation regrowth in abandoned agricultural plantations (in the Budjala region - northern DRC) that was mapped such as secondary forest, leading to disagreements of type 'PB-derived tree cover' — 'OB no tree cover' (see Figure 1.8).

1.6 Conclusion

This chapter developed an analysis framework which demonstrated how the pixel-based information could be accounted for at the object-level, in order to achieve a comparison between the two approaches. For the comparison analysis purpose, a PB-derived product was derived by upscaling the FACET pixels at the object level. The classification of the PB-derived objects was based on the tree cover thresholds that were used in the OB processing chain.

The agreement between PB-derived and OB products was very high, mainly due to the correspondence between 'PB primary forest' and 'OB no tree cover' classes. In turn, most of the disagreements between the two products were attributed to the rural complex areas (= 'OB non forest') that mainly corresponded to the secondary forest in the FACET data set. While the specific forest cover change dynamics occurring in this type of areas were mapped such as changes in the secondary forest domain, research is still need to clearly differentiate (young) secondary degraded forest from rural complex areas. Such a differentiation is challenging task since these land-cover types are very much mixed. To achieve this task, the OB processing chain could account for some additional spatial information such as the proximity to townships to distinguish between rural complex and secondary growth forest.

Additionally, the disagreements between the PB-derived and the OB classes also illustrated the need for an assessment of the FACET data set, which is for now the unique such spatially detailed information on the DRC forest cover and forest cover loss. Given the lack of national forest inventory data, this assessment could rely on field measurements or high spatial resolution imagery for selected locations. A class of particular interest for the FACET data assessment could be the secondary humid tropical forest and its associated forest cover losses. Within the secondary forest domain, the areas of interest for this

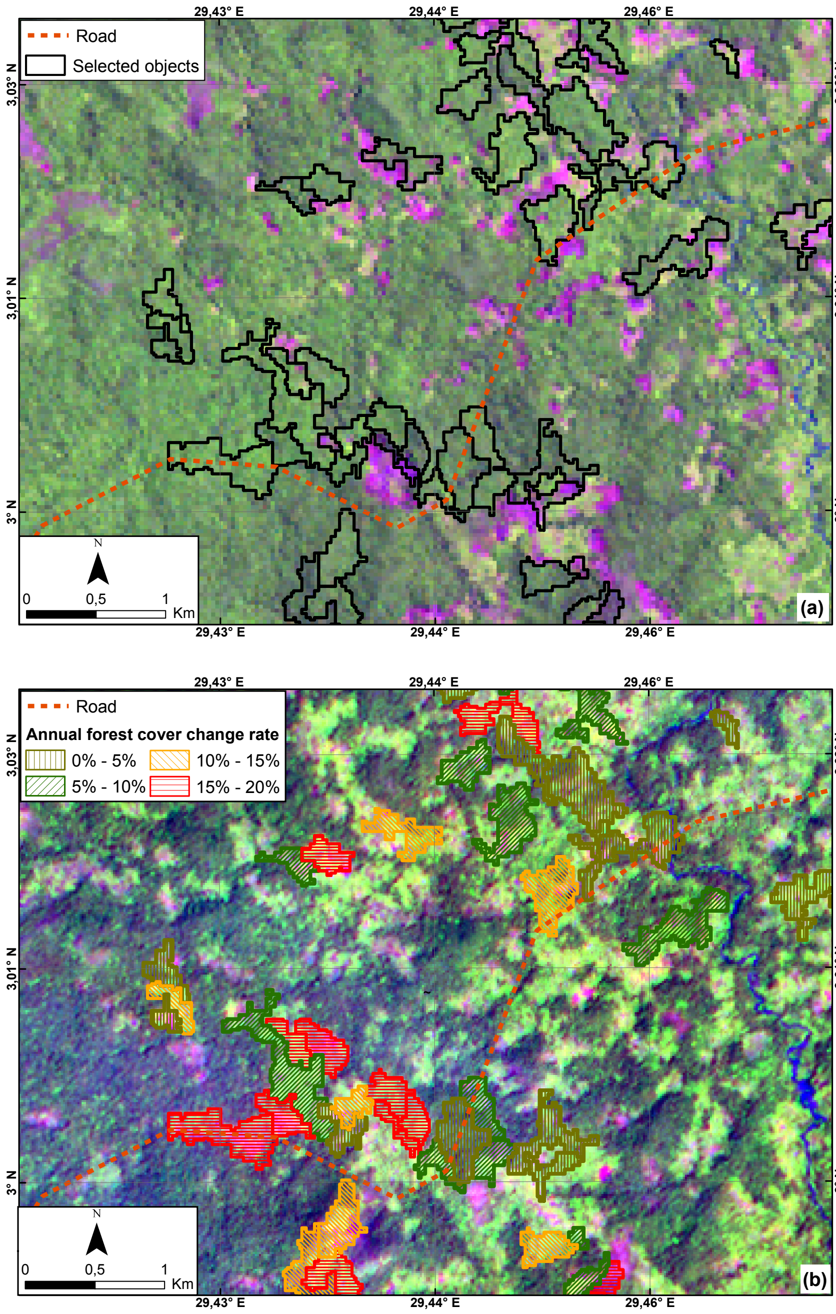
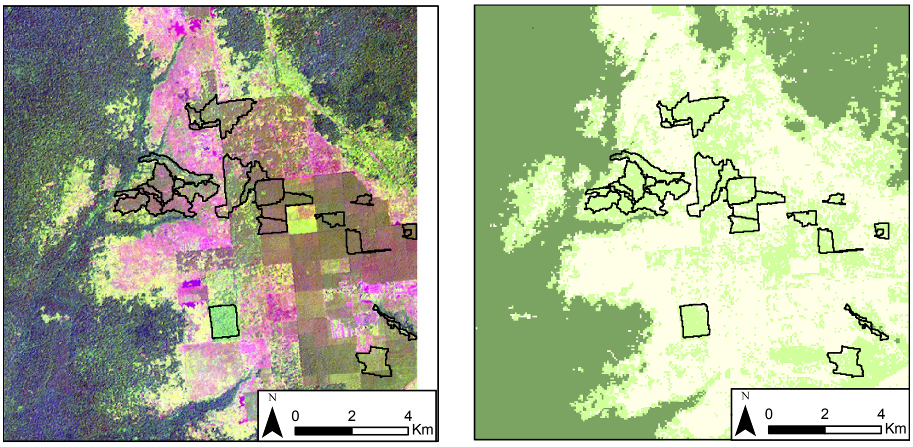


Figure 1.7: (a) Selected objects overlaid on the Landsat image 3°N;29.5°E of 05/05/2000. (b) Rates of forest cover change per object overlaid on the same Landsat image as in (a) on 03/12/2007. RGB=SWIR-NIR-Red.



Disagreement 'PB-derived tree cover' - 'OB no tree cover'. Landsat image 2,5°N;20°E of date 10/01/2001.

Figure 1.8: Vegetation regrowth in old agricultural plantations - Disagreements between OB and PB-derived products. Delineated objects are overlaid either on the Landsat image (RGB=SWIR-NIR-Red) or on the derived FACET map for year 2000 (PF in green, SF in light green, WO in brown, and NF in white).

assessment could be those corresponding to the OB rural complex areas.

The proposed method successfully combined information from pixel- and object-based approaches to obtain consolidated estimates of forest cover change. While some disagreements appeared, this kind of comparison contributes to the credibility and the transparency of forest cover change monitoring from Earth observation, today considered such as critical information for decision makers.

Chapter 2

Human population distribution modeling *

Abstract

Currently, there is a real need for a reliable assessment of the human population distribution in the Democratic Republic of the Congo which lacks the proper resources to support a large-scale census. A double-constrained-regression method (DCRM) was used to derive the population densities per land-cover type. Urban areas were found having the highest population density (about 4,400 inhabitants/km^2), followed by intensive and extensive agricultural areas (324 and 145 inhabitants/km^2), mountain forest (94), savanna (25), and other classes (2). Simulations were used to analyze discrepancies between census-derived and DCRM-predicted populations at the source zone scale. By comparison with freely available population data sets, the DCRM was found to poorly minimize these discrepancies while it was the only data set providing population densities' confidence intervals. Also, the DCRM output did not suffer from the abrupt changes in the population distribution such as observed for the other data sets used for comparison.

2.1 Introduction

Spatial models of human population distribution present many challenges. Typically, population modelers are faced with limited data availability, both in quantity and quality. As data are sourced mostly from national census campaigns, they tend to be inconsistent across both temporal and geographic scales. This is particularly true for African developing countries, which often lack the proper resources to support large-scale censuses (Mennis, 2009; Baker *et al.*, 2010). In best cases, population counts for these countries are made available at census units' scale. Most of the time, these information are, however, aggregated at the second or third country administrative level, which are territories

*Adapted from Kibambe Lubamba, J.-P., Bogaert, P., Nackoney R., J., and Defourny, P. (To be submitted) Assessing the human population distribution in African developing countries using a double-constrained-regression model. *Journal of Land Use Science.*

that may sometimes cover thousands of square kilometers (Elvidge *et al.*, 1997).

Some of the difficulties in modeling the human population distribution can be partially circumvented by the use of proxy data which can provide complementary and more spatially-detailed secondary information. In doing so, population distribution models mostly used geographic information systems (GIS) capabilities to allocate population counts over a regular grid, under assumptions related to a diverse array of ancillary data. These models therefore provide spatially-explicit data set of human population distribution.

Freely available and most relevant human population distribution data for African developing countries include the Gridded Population of the World (GPW), the Global Rural-Urban Mapping Project (GRUMP) database, LandScan, and the AfriPop data. All of these data sets were derived from variant dasymetric mapping approaches that are briefly described here below.

The latest version of the Gridded Population of the World (GPW-v3) was released in 2005. Input data are composed of individual countries' most recent census data describing population estimates by administrative units of varying resolutions. These data were transformed into a global 2.5 arc-minute grid or 5 km resolution using a simple areal weighting technique (Deichmann *et al.*, 2001). This technique is the baseline dasymetric mapping method which assumes a homogeneous distribution of human population within source zones and therefore oversimplifies the reality (Salvatore *et al.*, 2005; Langford, 2006; Baker *et al.*, 2010). This technique can in turn be useful for areas where no information is available to suggest a non-uniform distribution of the human population (Flowerdew and Green, 1992; Langford, 2006).

The GRUMP data set, mapped at 1 km spatial resolution, was based on a binary dasymetric method to allocate people into urban and rural areas using most recent censuses data (Salvatore *et al.*, 2005). The binary dasymetric method subdivided the source zones into populated and non populated subregions, and furthermore allocated source zone populations to the populated sub-regions only. The iterative process used to construct the GRUMP data set intended to match, when possible, the United Nations' estimated population growth rates. Urban and rural areas were explicitly distinguished, which represented a valuable improvement for global population distribution mapping.

The LandScan data set was mapped at 30 arc-second, or 1 km resolution, using 'smart' interpolation to model the global population distribution. Population occurrence by grid cell is based on weighted variables describing locations of roads, slopes and land-cover types, along with location and intensity of nighttime lights (Dobson *et al.*, 2000). In addition, very high resolution (1-5 meters) satellite imagery was incorporated into the modeling process to delineate urban extents and validate resulting population distribution (Dooley, 2005). Assumptions behind LandScan data are however not well documented (Salvatore *et al.*, 2005), which prevents any straightforward replication of the model using alternate data sets.

In the framework of the AfriPop project (`www.afripop.org`), Tatem *et al.* (2007) proposed a methodology for mapping both human populations and locations of settlements at high spatial resolution (100 m) for all African countries. Urban areas are delineated with a high degree of accuracy and population counts are allocated to these areas and to other land-cover types accordingly.

GPW-v3, GRUMP, LandScan, and Afripop share in common the fact that they adjust the population totals (Tatem *et al.*, 2011) so as there is no error in terms of population at the source zone scale. Indeed, LandScan adjusts its totals to match to match the estimates made by the Central Intelligence Agency (CIA) (CIA, 2010), while GPW-v3 and GRUMP adjust to the United Nations Population Division (UNPD) (UNDP, 2008). Population distribution data from Afripop area adjusted using population totals derived from intercensal growth rates (Hay *et al.*, 2005). For the case of the DRC, the adjustment at the source zone scale is problematic since these local estimates are projections from the 1984-census without any other reliability information. In addition, ancillary data used for these three population distribution data sets were chosen based on diverse assumptions. These were however rarely well established, somehow unclear, or a priori knowledge. Population distribution modelers accordingly used statistical techniques for quantifying the functional relationship between population counts within source zones and ancillary data (Langford *et al.*, 1991; Reibel and Agrawal, 2007; Liu *et al.*, 2008).

Proxy-data used to refine estimations of human population distribution include among others the transportation network (Dobson *et al.*, 2000; Reibel and Bufalino, 2005; Yue *et al.*, 2005), nighttime lights (Elvidge *et al.*, 1997; Sutton *et al.*, 1997; Small *et al.*, 2005; Briggs *et al.*, 2007), land-cover types (Dobson *et al.*, 2000; Reibel and Agrawal, 2007; Tatem *et al.*, 2007; Mennis, 2009; Baker *et al.*, 2010), and landscape metrics (Liu *et al.*, 2008). Meantime, various authors (Tian *et al.*, 2005; Tatem *et al.*, 2007; Lo, 2008; Mennis, 2009) considered land-cover data such as containing sufficient information to significantly improve the accuracy of population mapping. This improvement is especially related to the fact that land-cover data are usually derived using remote sensing technologies, which presents a viable and realistic alternative to obtain updated land-cover information to further estimate the human population distribution. This is particularly useful for African developing countries lacking resources for census campaigns, and countries covering large spatial extents and sometimes containing areas which are not easily accessible due to terrain or safety.

The relationship between contributing pixel counts of different land-cover types within source zones and corresponding population counts is most often achieved through variant of regression methods (Goodchild *et al.*, 1993; Yuan *et al.*, 1997; Langford, 2006; Baker *et al.*, 2010). Among these are constrained-regression methods that are used to force all regression coefficients to be non-negative (Moxey and Allanson, 1994), force estimated coefficients to satisfy the total population in the region of interest, or set possible negative coefficients to zero (Goodchild *et al.*, 1993). Regression coefficients can also be constrained to satisfy inequalities (Geweke, 1986). Goodchild *et al.* (1993) evenly evoked the introduction of constraints based on prior knowledge, such as a known population density in a particular sub-region.

The present chapter aims firstly at exploring the potential of a linear constrained - regression method to improve estimates of human population distribution in the Democratic Republic of the Congo (DRC). The method uses two specific constraints to estimate population densities per land-cover type, based-on population counts within source zones. These constraints are (1) the country's total population and (2) a prior knowledge about the percentage of people settled in rural areas. Secondly, the research demonstrates how information gained from both the land-cover map confusion matrix and simulated varying densities per land-cover class across source zones could be accounted for to assess the reliability of derived population densities. Results coming from the proposed method are compared to population estimates coming from GPW-v3, LandScan, and AfriPop data sets.

2.2 Data

The most recent DRC population census was conducted in 1984 and population data were aggregated to the Territory administrative level. This aggregation presents some major drawbacks. First, because most Territories are so large (Bafwasende, for example, stretches over 46,470 km^2), they prevent an accurate understanding of the population distribution dynamics at the national scale. More generally, such as suggested by Elvidge *et al.* (1997), population data aggregated at the administrative level conceal spatially-explicit details of population distribution at the finer scale. In addition, DRC's rural populations are crucial to understand population dynamics in the country. Indeed, rural areas are estimated to comprise about 70% of the country's total population (Herderschee *et al.*, 2012).

The data set used comprise a combination of census-derived populations at the Territory level and land-cover data for the year 2005. Details for these information sources are provided below.

2.2.1 DRC 2005 population data

De Saint Moulin (2006) updated the DRC population counts at the Territory level for the year 2005 based on the 1984 census. The author used different projection factors per Territory according to historical patterns of population growth (including restraints to growth resulting from recent wars). The mean annual population growth from 1984 to 2005 for the whole country was estimated to be approximately 3.16%. These census-derived population counts are used here as input for the model to spatially allocate the DRC total population

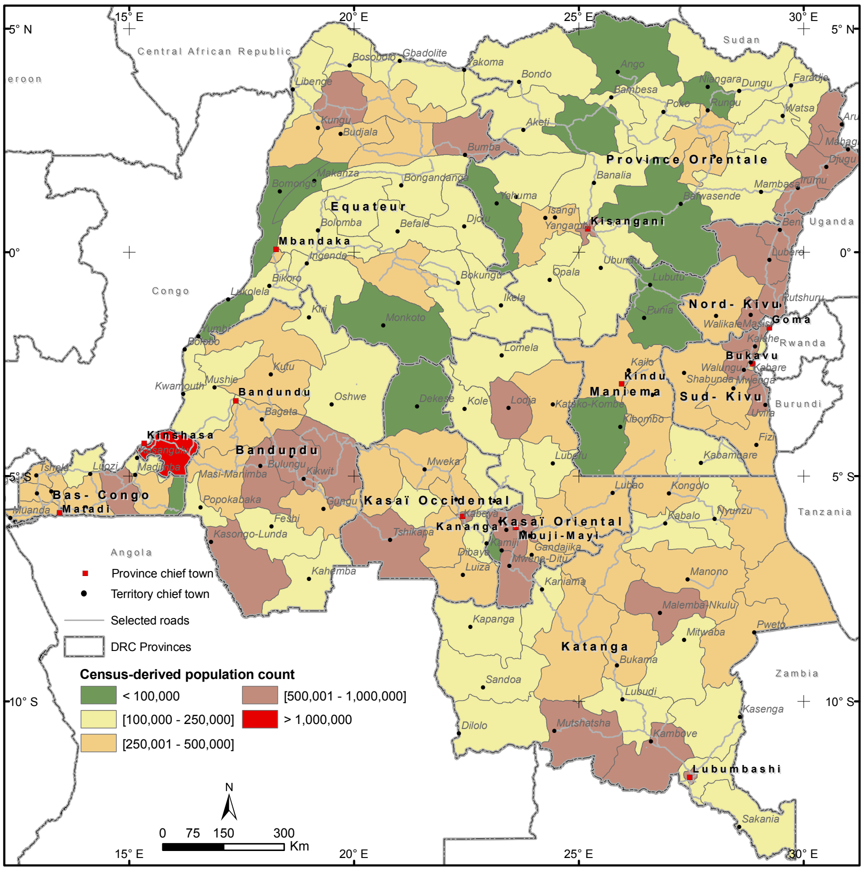


Figure 2.1: Census-derived population counts at the DRC Territory level.

across grid cells of 1-km resolution.

2.2.2 DRC 2005 land-cover data

The developed model employs land-cover classes as a proxy for the human population distribution. Data describing DRC's land-cover classes was obtained from Vancutsem *et al.* (2008). This data set was mapped at 1-km resolution and based on multi-temporal and multi-spectral analyses of SPOT VEGETATION images. The Table 2.1 summarizes the land-cover categories that are used. They are aggregated into six main classes, namely: urban areas, intensive agriculture, extensive agriculture, savanna, mountain forest, and a sixth class grouping together other land-cover types (e.g., young secondary forest, old secondary forest, and woodland). These were expected to contain similar very low population densities (this will be explicitly validated later on). The urban class was extracted from the Africover data set (Latham, 2001), which was mapped at a higher resolution (30 meters). The Africover data set has therefore a distinct advantage over the SPOT classification for delimiting urban areas.

Table 2.1: DRC land-cover classes and areas based on Vancutsem *et al.* (2008) and aggregated classes.

Class	Area (km^2)	%	**Class**	Area (km^2)	%
Aquatic grassland	1,863	0.08	Tree savannah	132,319	5.68
Urban	5,407	0.23	Edaphic forest	137,532	5.90
Mountain forest	6,602	0.28	Woodland (Miombo)	154,508	6.63
Agriculture	14,568	0.62	Old secondary forest	155,487	6.67
Submontane forest	32,638	1.40	Shrubland	171,760	7.37
Savanna woodland	45,893	1.97	Grassland	176,006	7.55
Rural complex	65,708	2.82	Steppe savanna-crops mosaic	314,780	13.51
Forest-savanna mosaic	84,682	3.64	Dense moist forest	703,643	30.20
Young secondary forest	126,168	5.42			
Aggregated classes					
Urban	5,407	0.23	Extensive agriculture	65,708	2.82
Intensive agriculture	14,568	0.62	Savannah	575,468	24.70
Mountain forest	39,240	1.68	Others	1,474,589	69.30

2.3 Methods

2.3.1 Double-constrained-regression

Assuming that the observed total population over a given area is mainly dictated by the land-cover types over this area, the corresponding average population density is computed as the simple ratio of population counts and land-cover type areas. The clear advantage of these density estimates obtained per land-cover type is that they can be used afterward to estimate populations over any other areas and at any other spatial scale, providing that the corresponding land-cover composition is known. Remotely sensed data are thus helpful with this respect, as they are able to deliver land-cover maps after image classification.

Deriving population density per land-cover type would be an easy task as long as census units are available over small areas, so that single land-cover types can be associated to each of them. For most practical studies, however, this proved to be impossible because available census data are most of the time aggregated at coarse scales. Moreover, density estimates are only meant to be average values, with expected local deviations due to numerous (and sometimes difficult to identify) factors. Discrepancies could result from spatially varying effects, whereas true errors could be inherent to the data themselves. E.g., land-cover maps obtained from remote sensing are not exempt of classification errors, whereas census data could be projections built from older data. It is therefore crucial to provide an interval of plausible population densities per land-cover type instead of single estimates without any attached reliability information. These issues can be circumvented using an explicit statistical modeling where errors from both image classification and census data are accounted for. The developed method explored this approach.

We denoted y_i as the observed population over the ith administrative unit among n units, and considered a set $\{\ell_1, \ldots, \ell_k\}$ of land-cover types, and $x_{j1}, \ldots, x_{jk}$ the corresponding areas (in km^2) covered by each land-cover type in this unit. By assuming that densities are spatially invariant, a simple linear regression model could thus be written as:

$$y_i = \underbrace{\beta_1 x_{i1} + \cdots + \beta_j x_{ij}}_{\widehat{y_i}} + \varepsilon_i \qquad \forall j = 1, \ldots, n \tag{2.1}$$

where β_j is the population density (inhabitants per km^2) associated with the land-cover class ℓ_j, and where ε_i refers to the corresponding error in terms of population. This model corresponds to a no-intercept model, in the sense that a null area (i.e., when $x_{ij}, \ldots, x_{ik}$ are null) necessarily yields a zero population (Yuan *et al.*, 1997; Langford, 2006; Reibel and Agrawal, 2007; Mennis, 2009). The estimation of the β_j parameters is straightforward in a classical ordinary least squares framework. Using standard matrix notations, the previous model and its solution are given by:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \Longrightarrow \quad \widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad \text{where } \mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nk} \end{pmatrix} \tag{2.2}$$

and a corresponding co-variance matrix $\Sigma_{\widehat{\beta}} = \mathbf{X}'\mathbf{X}^{-1}\sigma^2$. Accordingly, estimates $\widehat{\mathbf{y}}$ the population are given by $\widehat{\mathbf{y}} = \mathbf{X}\widehat{\boldsymbol{\beta}}$, with $\Sigma_{\widehat{\mathbf{y}}} = \mathbf{X}\Sigma_{\widehat{\beta}}\mathbf{X}'$.

Some land-cover types, such as urban areas and forests, are expected to naturally yield very different density estimates. Other close cover types like savanna woodlands, woodlands, and tree savanna could in turn be assumed to yield similar population densities. Useful grouping of these latter classes could therefore be justified and formally tested. On the opposite, it may appear that some originally intuitive classes (such as agriculture) are too coarse, in the sense that a finer definition (e.g., intensive *vs.* extensive agriculture) could yield significantly different density estimates.

Notwithstanding potential validity issues about the estimated densities (e.g., negative values), a major concern when applying this regression-based method is the adequacy between the observed total population $y_T = y_1 + \cdots + y_n$ and the estimated total population $\widehat{y}_T = \widehat{y}_1 + \cdots + \widehat{y}_n = \mathbf{1}'\widehat{\mathbf{y}}$. As linear regression only aims at minimizing the sum of the squared ε_i, there is no control over the estimated total population as resulting from the model (Goodchild *et al.*, 1993). This was circumvented in the developed method by introducing linear constraints using the Lagrangian formalism.

Considering that $SSE(\boldsymbol{\beta})$ represents the sum of squared errors ε_i^2's, is a convex functional of the set of β_j's , with

$$SSE(\boldsymbol{\beta}) = \mathbf{y}'\mathbf{y} - \mathbf{2}\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \tag{2.3}$$

The traditional solution for least square regression corresponding to the Expression 2.2 is obtained as:

$$\frac{\partial SSE(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + \mathbf{2}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \tag{2.4}$$

Considering the single linear constraint $\widehat{y}_T = y_T$, the objective function $\phi(\boldsymbol{\beta}, \lambda)$ to be minimized is given by:

$$\phi(\boldsymbol{\beta}, \lambda) = SSE(\boldsymbol{\beta}) + 2\lambda(\widehat{y}_T - y_T) \tag{2.5}$$

where λ is the Lagrangian multiplier. From Expression (2.1), it could be seen that

$$\widehat{y}_T = \sum_{i=1}^{n} \widehat{y}_i = \mathbf{1}'\mathbf{X}\boldsymbol{\beta} = \mathbf{s}'\boldsymbol{\beta} \tag{2.6}$$

with $\mathbf{s}' = (s_1, \ldots, s_k)$, where s_j is the total area (km^2) associated to the corresponding land-cover class ℓ_j as computed from:

$$s_j = \sum_{i=1}^{n} x_{ij} \quad \forall j = 1, \ldots, k \tag{2.7}$$

or stated in other words, $\mathbf{s}$ is the vector corresponding to the sum over the lines of $\mathbf{X}$. Substituting Expression (2.6) into (2.5) and taking the derivatives of Expression (2.5) both with respect to $\boldsymbol{\beta}$ and λ, the solution is given by:

$$\begin{cases} \dfrac{\partial \phi(\boldsymbol{\beta}, \lambda)}{\partial \boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{y} + \mathbf{2X'X}\boldsymbol{\beta} + \mathbf{2}\lambda\mathbf{s} = \mathbf{0} \\ \\ \dfrac{\partial \phi(\boldsymbol{\beta}, \lambda)}{\partial \lambda} = 2(\mathbf{s}'\boldsymbol{\beta} - y_T) = 0 \end{cases} \iff \begin{pmatrix} \mathbf{X'X} & \mathbf{s} \\ \mathbf{s}' & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ y_T \end{pmatrix} \tag{2.8}$$

Though this illustrates the use of a single constraint on the total population, it is possible to account for multiple linear constraints by working along the same line. As a specific example for DRC, 30% of the total population is settled in urban areas, whereas the remaining 70% is spread over the other land-cover types. Accordingly, the new solution is given by:

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{s}_a & \mathbf{s}_b \\ \mathbf{s}_a' & 0 & 0 \\ \mathbf{s}_b' & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \lambda_a \\ \lambda_b \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{y} \\ (0.3)y_T \\ (0.7)y_T \end{pmatrix} \tag{2.9}$$

with $\mathbf{s}_a' = (s_1, 0, \ldots, 0)$ and $\mathbf{s}_b' = (0, s_2, \ldots, s_k)$, where s_1 denotes the total urban area, respectively, whereas $s_2, \ldots, s_k$ refer to total areas for the various other classes. Similarly, the same procedure could be applied to account for specific information about other classes, or over specific aggregated areas (Langford, 2006; Mennis, 2009).

As a final word, it is worth noting that for this regression approach, the null model corresponding to the absence of land-cover effect on the population density is written as:

$$y_i = (x_{i1} + \cdots + x_{ij})\beta + \varepsilon_i \quad \forall i = 1, \ldots, n \quad \iff \quad \mathbf{y} = \mathbf{s}\beta + \boldsymbol{\varepsilon} \tag{2.10}$$

i.e., there was a single population density β to be computed over the whole country. Solving for β in a classical unconstrained least squares framework yields the non intuitive result:

$$\widehat{\beta} = (\mathbf{s}'\mathbf{s})^{-1}\mathbf{s}'\mathbf{y} = \frac{\sum_{i=1}^{n} s_i y_i}{\sum_{i=1}^{n} s_i^2} \tag{2.11}$$

On the opposite, using the additional constraint $y_T = \widehat{y}_T$ where $\widehat{y}_T = \mathbf{1}'\mathbf{s}\beta$ directly forced the only possible result:

$$\widehat{\beta} = \frac{y_T}{\mathbf{1}'\mathbf{s}} = \frac{y_T}{\sum_{i=1}^{n} s_i} \tag{2.12}$$

i.e., the single density is computed as the ratio of the total country population divided by the total country area, as expected. More generally, this also highlights the influence that constraints may have on density estimates in the case of less trivial models.

2.3.2 Discrepancy analysis

Census-derived and predicted populations per administrative unit were compared to highlight under and over-estimations. Possible sources of errors were related either to the reliability of the data themselves (i.e., the census counts per administrative unit and the land-cover map), or to underlying assumptions of the developed method.

DRC census-derived populations might seriously depart from true values as they are merely projections from previous years. On the other side, classification errors are expected to be present in the land-cover map, which might affect the estimated population densities. Finally, it was assumed that population densities per land-cover types are identical for all administrative units, while they could be varying over the whole country. Next paragraphs illustrate how the proposed method analyzes these potential sources of errors.

Thanks to the availability of a confusion matrix (CM) for the DRC land-cover map (Vancutsem *et al.*, 2008), it was possible to evaluate if the discrepancies between census-derived and predicted population counts could be attributed to the classification errors. For this task, the CM was adapted in order to fit the 6 chosen land-cover classes (Table 2.2). It was assumed that urban and mountain forest were always correctly classified, while extensive and intensive agriculture (that were not distinct in the original matrix) exhibited the same confusion errors with the other classes. A simulation procedure was used to assess the effects of these classification errors. According to the values given in the confusion matrix, pixels were randomly reallocated to a given land-cover class on a pixel basis over the whole map. Each simulated land-cover map thus includes classification errors in the same global proportions as those mentioned in the confusion matrix. This led to both different $\mathbf{X}_i$, matrices, different densities, and population estimates. The effect of classification errors was highlighted by comparing predicted population counts using the double-constrained-regression method (DCRM) to those obtained using simulated land-cover maps.

The second possible source of errors is an irrelevant assumption of constant densities per land-cover type over the whole country, as these values could naturally vary from places to places. However, no precise information is available about the distribution of population density over the country. The option here was thus to assume that these distributions correspond to those obtained from regression. At the spatial resolution of the administrative units, it is indeed unlikely that the range of possible density values over these units would exceed these already quite large intervals. A set of 143 density vectors $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{143}$ (i.e., one different set per administrative unit) was independently

Table 2.2: Adapted confusion matrix of the 2005 DRC land-cover map.

	Urban	Int. agr.	Ext. agr.	Mountain f.	Sav.	Others
Urban	1	0	0	0	0	0
Int. agr.	0	0.978	0.007	0	0.001	0.014
Ext. agr.	0	0.007	0.978	0	0.001	0.014
Mountain f.	0	0	0	1	0	0
Savannah	0	0.064	0.064	0	0.500	0.370
Others	0	0.011	0.011	0	0.015	0.964

simulated based on multivariate Gaussian hypothesis i.e., $\boldsymbol{\beta}_j \sim N\left(\widehat{\boldsymbol{\beta}}, \Sigma_{\widehat{\boldsymbol{\beta}}}\right)$ to assess the effect of different densities per administrative units. For all administrative units, corresponding population estimates using spatially varying densities were thus compared to those previously obtained using single density estimates.

2.3.3 Comparison of population data sets

The DCRM output was compared to most relevant and freely available population data sets for the DRC. These were namely the GPW-v3, LandScan, and Afripop. This comparison was carried out at two different stages. First, the sum of squared errors (SSE) between census-derived and predicted population counts was computed for each of these population distribution data sets, DCRM included. Second, a spatial comparison of these data sets was also performed to highlight possible abrupt and unexpected changes in the resulting population distribution maps.

2.4 Results

Estimated human population densities per land-cover type from the simple- and double-constrained-regression models show little differences at the 95% of confidence interval (Table 2.3). The percentage of estimated total population over urban areas is equal to 30.25% for the single-constrained-model which tends to prove that the reported 30%-70% proportions are indeed in close agreement with reality.

Table 2.3: DRC unconstrained and constrained regression estimates, along with their corresponding 95% confidence intervals.

	No constraint			**One constraint**			**Two constraints**		
Class	Est.	2.5 %	97.5 %	Est.	2.5%	97.5 %	Est.	2.5%	97.5 %
Urban	4,136	3,271	5,000	4,429	3,562	5,296	4,393	3,526	5,260
Int. agr.	314.97	126.16	503.78	323.37	134.08	512.66	324.30	135.01	513.59
Ext. agr.	130.98	59.27	202.68	146.01	74.13	217.90	145.00	74.11	217.89
Mountain f.	85.78	40.81	130.75	93.49	48.41	138.58	93.57	48.48	138.66
Savannah	21.41	13.40	29.41	25.31	17.28	33.34	25.42	17.39	33.44
Others	-0.45	-4.12	3.22	1.87	-1.81	5.55	1.90	-1.78	5.59

Results from the DCRM output (Figure 2.2 and Table 2.2) show that urban areas are still have the highest population density (about 4,400 inhabitants / km^2), followed in a decreasing order by intensive agricultural areas (324 inhabitants / km^2), extensive agricultural areas (145 inhabitants / km^2), mountain forest (94 inhabitants / km^2), savanna (25 inhabitants / km^2), and other classes (2 inhabitants / km^2). Some particular administrative units, which were not included into the modeling process, are hatched on the Figure 2.2. These are the DRC capital Kinshasa that is deemed to be a particular case of urban area densely population over a restricted area, along with the coastal city of Moanda, which exhibited a census-derived population clearly inconsistent with its associated land-cover classes.

Discrepancies between census-derived and DCRM-predicted populations at the source zone scale (Figure 2.3a) are analyzed using simulations. Thus, the effects of classification errors and those resulting from considering a uniform population density per land-cover type for the whole country are evaluated. The Figure 2.3c and the Table 2.4 show that classification errors have a very limited impact on both density and population estimates, so these estimates could be considered as robust with respect to this source of error, at least for the present study.

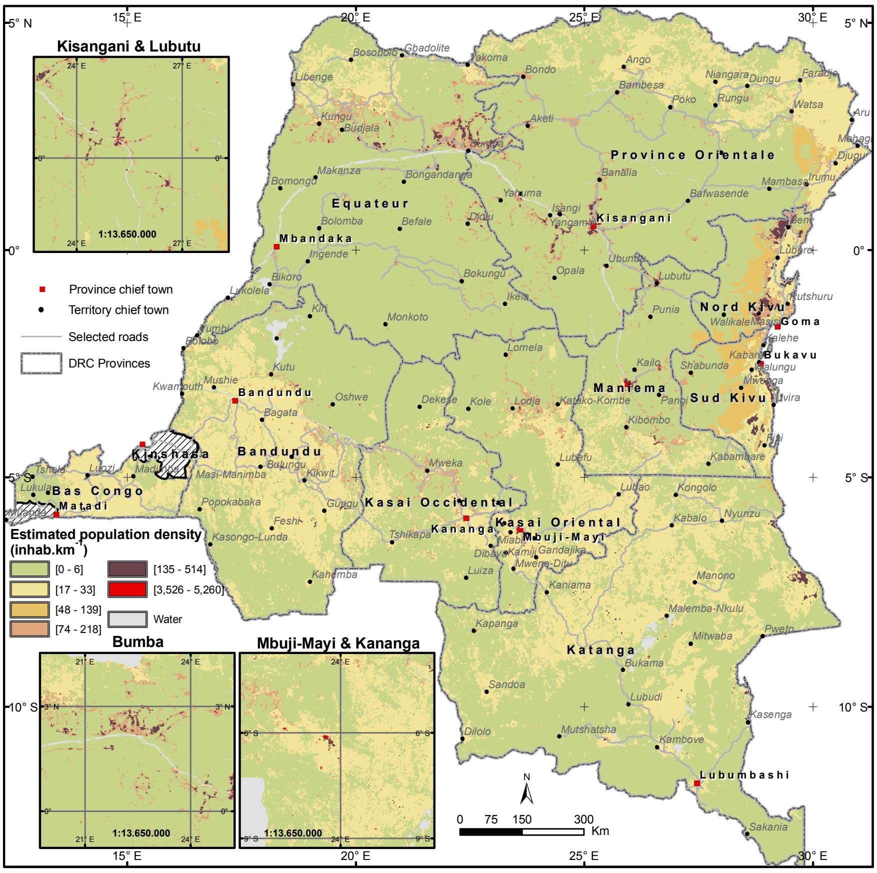


Figure 2.2: Human population distribution for the DRC from the DCRM.

Table 2.4: Analysis of residuals' variances.

	Variance of		
Class	Residuals	Spatially varying residuals	Residuals using the CM
Urban	451.9	451.9	12.5
Int. agr.	98.7	98.1	11.8
Ext. agr.	37.5	30.6	3.2
Mountain forest	1.9	1.7	0.1
Savannah	4.2	3.7	1.2
Others	23.5	22.8	0.6
Standard deviation	2.6e+005	4.5e+004	6.6e+003

From Figure 2.3b and Table 2.4, it can be seen that spatially varying densities have indeed an impact on the result, but this variability is far from explaining what is really observed.

The comparison between the DCRM output and LandScan, Afripop, and GPW-v3 data was conducted to compare (1) the census-derived and predicted total populations within source zones, and (2) the patterns of the population distribution. LandScan was found as the data set minimizing the differences between predicted and census-derived populations at the source zone scale (SSE=914,852), followed by Afripop (SSE=1,037,558), and GPW - v3 (SSE=1,364,888). The DCRM poorly achieved this goal (SSE=3,073,349), mainly due to the fact that the method was not designed to match local estimates for which there is no guarantee of reliability.

Three sampled locations were used to compare the patterns of population distribution such as resulting from GPW-v3, LandScan, Afripop, and DCRM data sets. These locations are (1) eastern DRC (Figure 2.4a), (2) the region around Bumba (Figure 2.4b), and (3) central DRC (Figure 2.4c). The eastern DRC is a region known to have high population densities, and it was observed that LandScan, Afripop, and GPW-v3 suffered from abrupt changes in the population distribution, particularly for areas located between contiguous source zones. This was mainly due to two factors. The first one is incumbent to the simple areal weighting technique that homogeneously allocates the total population within source zones, particularly for the GPW-v3 data set. The second factor relies on the proxy-data used. Indeed, LandScan and Afripop were found as maximizing the allocation of population counts at locations corresponding to ancillary data used, in order to model the population distribution. Remaining population was evenly allocated within the source zones' areas for which the proportion of ancillary data was approximately nil. DCRM data was in turn found as lacking these abrupt and unexpected changes in the population distribution.

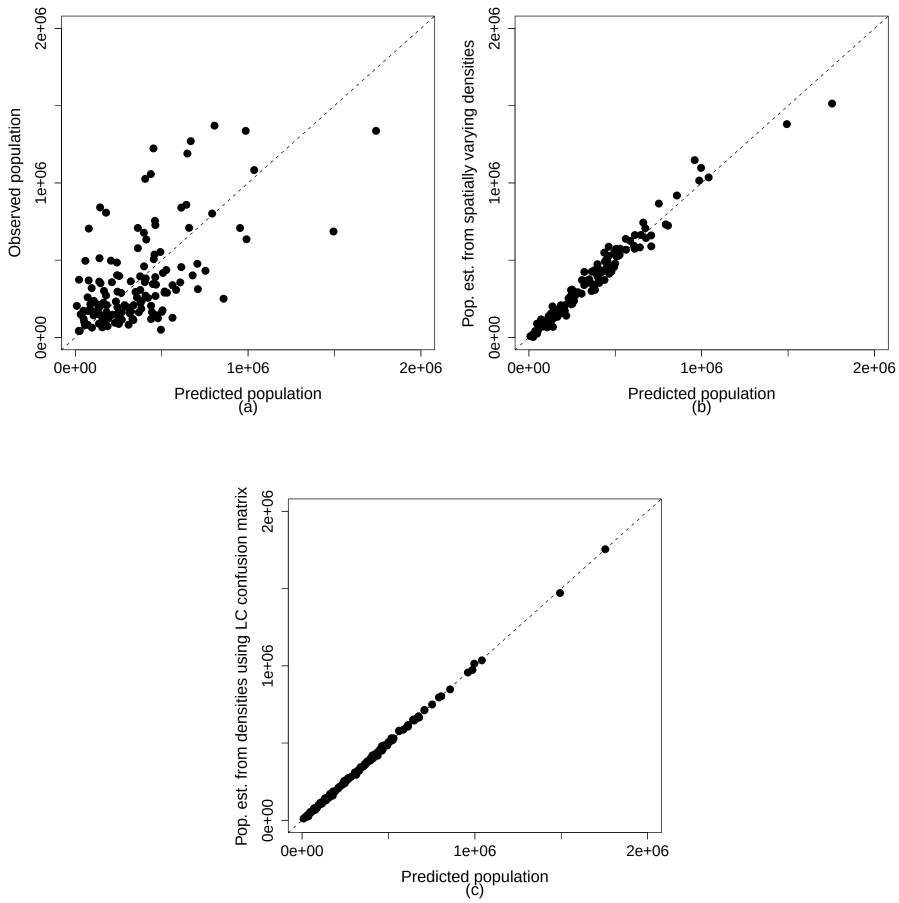


Figure 2.3: DCRM-predicted population at the source zone scale *vs.* (a) observed population, (b) population estimates using the land-cover confusion matrix, and (c) population estimates from spatially varying densities.

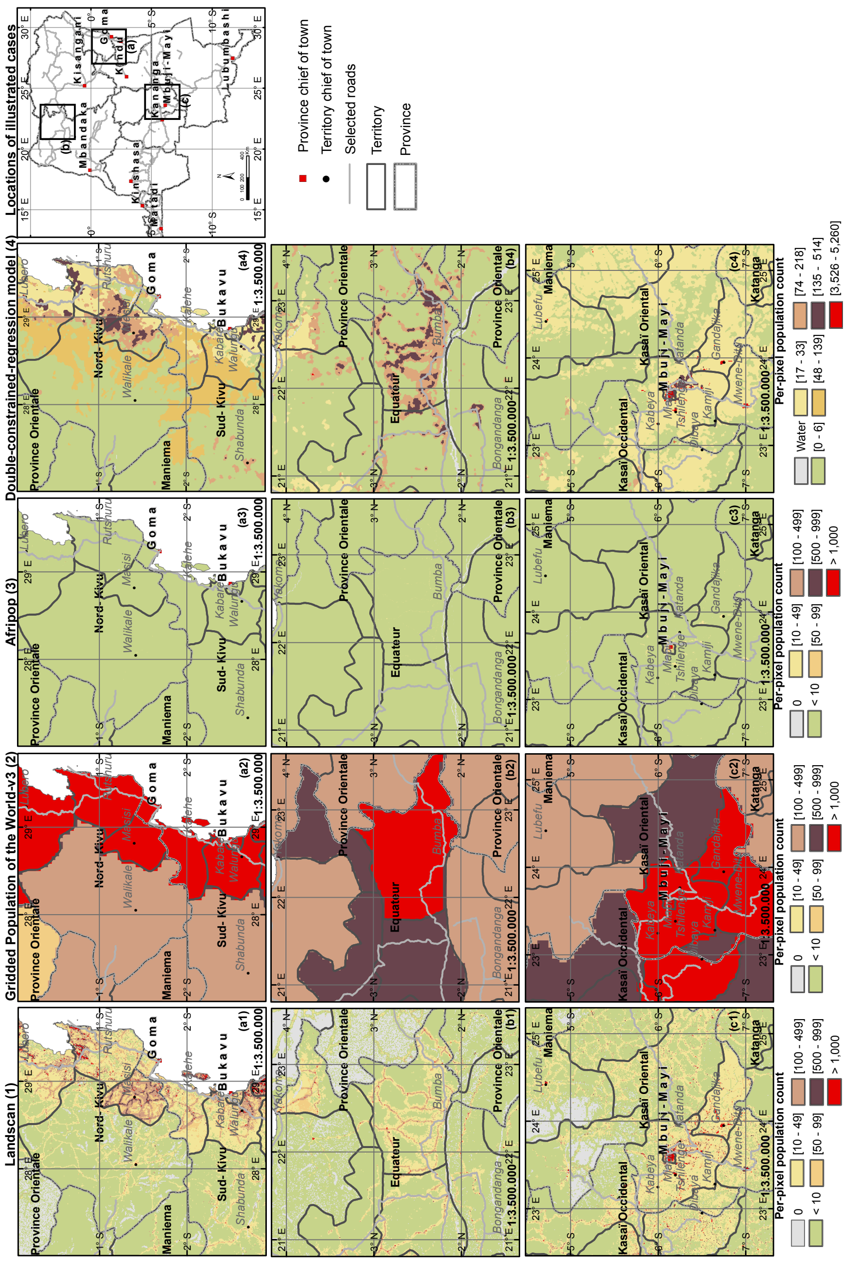


Figure 2.4: Spatial comparison of population data sets for three sampled locations: (a) eastern DRC, (b) Bumba, and (c) central DRC from (1) LandScan, (2) the GPW-v3, (3) Afripop, and (4) the DCRM.

The region around Bumba is known for its intensive agricultural activity. Despite of the Afripop finer resolution, this model did not map these rural areas' populations such as observed on both LandScan and DCRM outputs. The analysis of the central DRC region for its part raised the evidence of the finer delineation of human settlements achieved by both Afripop and LandScan data. These urbanized areas are not depicted by GPW-v3 while they are poorly mapped from DCRM data.

2.5 Discussion

Sound and reliable estimation of the human population distribution in DRC is an important prerequisite for a wide array of applications. Forest cover change modeling, vulnerability assessment, land-use planning, environmental projects, and others management issues cannot be handled efficiently without integrating the human population distribution component.

This chapter proposed a methodology to estimate the population densities by land-cover class using a double-constrained-linear regression model. These population densities further allowed estimating the population distribution in a spatially-explicit manner. The model used aggregated land-cover classes as a basis for predicting the population distribution. The modeling approach confirmed that constraining the total population notably reduced the bias between census-derived and predicted populations, while it was also found that estimated densities were on par with general assumptions about the DRC human population distribution. For example, the model depicted a medium density (about 94 inhabitants/km^2) for mountain forests, in the quite populated regions of North and South Kivu. This is a DRC specific case while mountainous ecosystems are elsewhere associated to very low human population densities.

The constrained-regression methodology was applied both on the 17 initial classes as well as on a grouping into 6 classes. The 17 original land-cover classes were expected to cover a wide range of population densities, spreading from several thousands inhabitants/km^2 for urban areas to few inhabitants/km^2 for forests. Using a formal F-test, grouping from 17 to 6 classes was statistically acceptable (p_v=0.15). Pooling further intensive and extensive agriculture classes would be statistically appropriate but the dynamics observed for these two classes were expected to impact differently the human population distribution. In fact, the extensive agriculture class mainly developed along the road network corresponds to human activity zones, while the intensive agriculture also refers to cities and villages not depicted due to the sensor spatial resolution (Vancutsem *et al.*, 2008). This latter class mainly surrounds cities such as Bumba, Kisangani, Kindu, Beni, and Masisi, which are quite populated places. For these reasons, it was thus decided to keep these two classes as distinct.

The observed gap between census-derived and DCRM-based populations at the source zone scale led to simulate spatially varying population densities, such as suggested by previous studies (Lo, 2008). The model showed that allowing estimated densities to vary over space improved the predicted populations at locale scale. It remained however doubtful that spatially varying densities could explain observed discrepancies between census-derived and DCRM-derived population counts. These discrepancies were also assessed against the classification errors that are inherent to land-cover maps. The proposed method thus suggested a framework to account for the land-cover map confusion matrix in order to evaluate the robustness of predicted population densities against the classification errors. While the corresponding errors were quite limited and were not expected to notably change the estimations for the specific DRC case study, it might prove to be useful to account for them in other circumstances.

Based on the previous findings, as both spatially varying densities and classification errors did not allow to fully explaining the discrepancies between observed and DCRM-derived populations, the reliability of census-derived population on hand were therefore questionable. In fact, 2005 DRC population used in this paper were projections from the 1984 DRC census which could result in severe under/over estimations compared to true observed populations. Unfortunately, the lack of reliable validation information on population at the DRC national scale did not allow investigating further this aspect in the framework of this chapter. Further investigation is also needed to reliably characterize the original 1984 census data (e.g., number and size of enumeration units, census strategy, census error) and estimate the error associated to the census-derived population projections obtained from De Saint Moulin (2006).

The comparison of LandScan, GPW-v3, Afripop, and DCRM data sets showed that LandScan minimized the discrepancies between census-derived and predicted total populations at the source zone scale, followed by Afripop and GPW-v3. While the DCRM poorly performed the goal to meet source zones' population totals, this latter data set provided, in turn, intervals of plausible human population densities. These intervals are of paramount importance considering the lower accuracy of available census-derived population counts in source zones for the DRC. Also, the different possible density values for each land-cover class can be used to simulate a large number of population maps from which it is possible to select the ones minimizing the differences at the source zone scale. Other population distribution models used for the comparison analysis were adjusted to match population totals at the source zone scale and only provided single estimates of population per pixel without any attached reliability information. However, DCRM densities intervals were quite large and could be refined by conducting field-based validation over small areas. This was not carried out in the present research due to time and cost constraints.

The spatial comparison of the population distribution data sets highlighted that abrupt and unexpected changes in population estimates occurred at both Territory and Provinces boundaries for GPW-v3, LandScan, and Afripop. Because these administrative boundaries are most of the time arbitrary limits (Crampton, 2004), they cannot explain these sudden changes in the population distribution (Monmonier, 1996; Mennis, 2009). These abrupt changes also illustrated that for LandScan, GPW-v3, and Afripop data sets, the requirement to match population totals in source zones biased the population distribution, especially for locations nearest to administrative boundaries. The DCRM overcame these unexpected changes in the population distribution, while the model is still tributary to the accuracy and the spatial resolution of the land-cover map used. The spatial resolution effect was particularly highlighted in the coarser mapping of human settlements compared to Afripop data. This pleads for updated and validated land-cover maps, at higher spatial resolutions to improve the detection of human population distribution patterns (Tatem *et al.*, 2007). These improved land-cover maps should not be focused only on enhancing the delineation of human settlements that are quite sparse in DRC rural areas, and somehow difficult to detect from satellite imagery of coarse resolution. Better land-cover maps should also refine the depiction of other land-cover types particularly those referring to human activities in rural areas.

It was also observed that LandScan data was particularly tributary of the DRC road network, which was used among a set of other proxy-variables to obtain the population distribution. Blaes (2009) recalled that while spatial data for primary roads in DRC are being updated and becoming increasingly available, information on secondary roads are, in turn, still scarce and most the time unavailable or outdated. The use of road-proximity as ancillary data to model the DRC human population distribution should therefore be done with caution.

2.6 Conclusion

The model developed in this chapter is simple to implement, easy to interpret, and could be used straightforwardly for alternative data sets or for other African developing countries. The model is also flexible enough to account for any possible additional constraint that one may have at hand about the total population for given land-cover classes, or even over given regions. The framework developed to account for the land-cover map confusion matrix and simulated spatially-varying densities proved to be useful to assess the discrepancies between census-derived and predicted population counts.

The comparison results with most relevant and freely available population data sets highlighted that DCRM poorly achieved the goal of minimizing differences between census-derived and predicted populations at the DRC Territory level. The developed method, however, provided confidence intervals for estimated population densities unlike the other data sets (including GPW-v3, LandScan, and Afripop) used for comparison. Also, accounting for the fact that there is a likely reliability issue for the census-derived population counts themselves, DCRM-derived population densities could be considered as reasonable and meaningful, according to what is known about the dynamics of the human population distribution in the DRC.

The spatial comparison of the DCRM output with these freely available population data sets showed that the DCRM output did not exhibit sudden and unexpected changes in the population distribution. The DCRM was, however, still tributary of the land-cover map spatial resolution and accuracy. This highlighted the benefit of using remote sensed images acquired at finer resolution to deliver improved land-cover maps, in order to delineate more accurately human settlements, as well as other land-cover types corresponding to human activity zones in rural areas. Remote sensing is thus helpful for population distribution mapping, particularly for a country such as the DRC which lack the proper resources to support national census campaigns.

Chapter 3

Accessibility modeling *

Abstract

Accessibility is a key driving factor for economic development, social welfare, resources management, and land use planning. In many studies, modeling accessibility relies on proxy variables such as estimated travel time to selected destinations. In developing countries, estimating the travel time is hindered by scarce information about the transportation network, making it necessary to take into account *off-network* travel coupled with considerations of multimodal options available within the existing network. This chapter proposes such a hybrid approach that computes the travel time to selected destinations by optimizing together a fully modeled multimodal network and *off-network travel*. The model was applied in a region around Kisangani located in northeastern Democratic Republic of the Congo. Travel times to Kisangani from the hybrid approach were found to be in close agreement with field-based information ($R^2 = 0.98$). The developed approach also proved to better support real-world transportation constraints (such as transfer points between travel modes or barriers) than cost-distance-based travel-time modeling. Demonstration results from the hybrid approach highlight the potential for impact assessment of road construction or rehabilitation, development of secondary towns or markets, and for land use planning in general.

3.1 Introduction

Today, accessibility is a key driving factor for economic development, social welfare, resources management, services access, and land use planning. At the same time, geographic information system (GIS) technologies made available to a wide community have the capability to compute accessibility maps for a large variety of studies. These studies include land-use and land-cover changes (Lambin *et al.*, 2006), health care programs (Martin *et al.*, 2002; Tanser *et al.*, 2006; McGrail and Humphreys, 2009; Moisi *et al.*, 2010), transport strate-

*adapted from Kibambe Lubamba, J.-P., Radoux, J. and Defourny P. 2013. Multimodal accessibility modeling from coarse transportation networks in Africa. *International Journal of Geographical Information Science*, 27 (5): 1005 - 1022.

gies (O'Sullivan *et al.*, 2000; Geurs and Van Wee, 2004), infrastructure building (Gutiérrez and Urbano, 1996; Gutiérrez *et al.*, 1996), and environmental economics (Brainard, 1999), among others.

Accessibility, therefore, comprises a variety of spatial, socioeconomic, and temporal aspects and accordingly can be quite complex to model (Geurs and van Eck, 2001; Geurs and Van Wee, 2004). Consequently, several studies considered the travel time to selected destinations as an accessibility proxy variable (Brainard *et al.*, 1997; O'Sullivan *et al.*, 2000; Liu and Zhu, 2004a,b; Castella *et al.*, 2005; Vandenbulcke *et al.*, 2009; Moisi *et al.*, 2010).

However, estimating the travel time in most of the developing countries is hindered by the scarcity or unavailability of the necessary data to inform input parameters (Perry and Gesler, 2000; Tanser *et al.*, 2006). Such data may describe certain characteristics of the transportation network (travel speed by road type, road condition, location of railway stations, timetable information for public transportation, multimodal considerations, etc.) and the type of destination (number of available services at the destination, competing aspects between available destinations, etc.).

Estimating travel time can be straightforwardly achieved using GIS, assuming input data are readily available. In many developing countries, data for primary roads are being updated and increasingly made available. However, information on secondary roads are still scarce and most of the time unavailable, preventing the implementation of reliable GIS-based travel-time models (Perry and Gesler, 2000). The frequently poor condition of both primary and secondary road networks in developing countries has had two major consequences: (1) the use of *off-road* travel and (2) the development of alternative transportation systems such as the river network. People therefore walk from villages (*off-road* travel) to the least effort-requiring road or river network transfer point, where means of transportation (car, motorbike, boat, etc.) allow further route to the destination.

According to Tanser *et al.* (2006), *off-road* travel should be considered when modeling accessibility in developing countries. Cost-distance-based travel-time modeling has been used widely to tackle issues coming from *off-road* travel and provided reliable approximations (Brainard *et al.*, 1997; Brainard, 1999; Nelson, 2000; Watts *et al.*, 2003; Verburg *et al.*, 2004b; Tanser *et al.*, 2006; Nelson, 2008). The cost-distance function uses a friction surface to calculate the least cumulative cost from each cell to the nearest destination for a given area.

One of the major weaknesses of cost-distance-based travel-time modeling is an assumption that states that during travel through the friction surface, network segments can be reached at any location on the transportation network, and not only at specified stops. However, many transportation axes (such as highways, railroad systems, or rivers) have specific connectivity and are reached in precise locations (such as highway entrances, train stops, harbors, or ferries locations). Certain exploratory studies tried to overcome the limitations of cost-distance-based travel-time modeling by implementing multimodal travel-cost modeling (Juliao, 1999). However, multimodal models are not designed for managing *off-network* travel.

As mentioned above, because of the poor condition of roads in developing countries, alternative transportation systems (such as navigable rivers) are often used to reach a destination. In these cases, potential roads–rivers interchanges should be captured in a single multimodal transportation system when modeling travel time.

GIS-based network analysis offers efficient tools for managing multimodal networks and provides outputs such as the optimal route (with or without barriers) and isochrones (Liu and Zhu, 2004a). Some accessibility models are particularly focused on origin — destination (OD) travel-time estimations to provide accessibility measures (de Jong and van Eck, 1996; Liu and Zhu, 2004a; Vandenbulcke *et al.*, 2009). It should be noted, however, that these OD models do not provide accurate travel times for the whole study area if all origin locations are not integrated into the modeling process. In fact, villages and small settlements (normally considered as origin points for travel-time modeling in developing countries) are often inaccurately located and even unreported in geographic databases (Blaes, 2009). For these regions, the most suitable approach for travel-time modeling should rely on estimating the travel time from *any* location of the study area to selected destinations. This approach results in the generation of a travel-time surface.

According to Brainard *et al.* (1997) and Martin *et al.* (2002), many accessibility models have resorted to isochrones for generating travel-time surfaces. O'Sullivan *et al.* (2000) used a vector-based approach to develop isochrone maps for Glasgow (Scotland). The authors developed a comprehensive modeling that proved to be sufficiently robust to use the isochrones for a constrained accessibility analysis. However, such a modeling approach requires a large amount of data and is quite complex to implement at a national scale. Most importantly, this approach needs to explicitly identify both origin and destination points and cannot accommodate situations where origin points are located anywhere in the study area. Also, accessibility values between network segments from resulting isochrone surfaces remain questionable in a coarser street grid region.

Brainard *et al.* (1997) used an integrated vector — raster model to tackle the problems coming from the use of isochrones in areas between network segments. After calculating the travel time through the road network from origin locations to the nearest destination, they used three approaches for computing travel time for areas away from network segments. The first approach was based on the strong assumption that areas located farther from roads do not have reduced access and, subsequently, their accessibility values were computed using a nearest neighbor algorithm. The major criticism of that approach, apart from its oversimplification, was that underlying friction surfaces between network segments — which can restrain or facilitate the transport flows — were ignored. However, this approach could be useful in most developed countries where road proximity could be computed simply by creating corridors along the network, which can be assumed as being of uniform density and quality in all directions (Gutiérrez and Urbano, 1996). The second approach was an improved variant of the previous. Gaps between roads were filled by incrementing the nearest cells' values with the same amount of time that was assumed to be a typical walking time to traverse the grid cell. The third approach was a hybrid between the two others: cells within 1.5 km of known roads were filled with the nearest neighbor algorithm and remaining cells were assigned an accessibility value using the second approach.

The main problem with the method developed by Brainard *et al.* (1997) resides in the use of Euclidean distances to primarily assign the network time values to neighbor cells. This assignment can create artifacts in at least three cases: (1) the closest road is not the fastest route to reach a given destination, (2) the shortest path (nearest neighbor) is not the fastest one due to potential barriers (e.g., transfers from *off-road* travel to main network only feasible at specified locations), and (3) the nearest neighbor is not the fastest route due to an underlying friction surface.

As a consequence, isochrone generation methods developed by O'Sullivan *et al.* (2000) and Brainard *et al.* (1997) do not explicitly provide a suitable solution for areas located between roads of different qualities that allow different travel speeds. These methods will then tend to induce certain discrepancies and artifacts (artificial jumps, sudden falls, or other protrusions) in estimated travel time. These discrepancies will be exacerbated for areas located farthest away from any network segment where the transportation network is coarse, particularly in African developing countries, such as the Democratic Republic of the Congo (DRC) (see Table 3.1).

Table 3.1: Road indicators in 2010 for some of the DRC Provinces compared to selected developed countries, adapted from Iaych (2010).

DRC/developed country	Pop. (2005) ($\times 10^6$ inh.)	Area ($\times 10^3$ km^2)	Road density (km/km^2)
Bas-Congo / Croatia	3.35 / 4.44	54 / 56	0.14 / 0.52
Sud-Kivu / Ireland	4.68 / 4.11	64 / 69	0.05 / 1.37
Bandundu / Italy	6.39 / 58.46	297 / 294	0.11 / 1.62
Equateur / Sweden	6.31 / 9.01	402 / 410	0.06 / 1.28
Pr. Orientale / Spain	6.97 / 43.04	497 / 499	0.05 / 1.32

Thereby, more realistic travel-time modeling in developing countries appears a challenging task that should respond to specific requirements coming from:

- the scarcity of the transportation system and its poor condition;
- the use of alternative routes (such as rivers and *off-roads* travels) for supplementing the road network;

- the possible restriction of transfers between network components (such as roads—rivers interchanges) which can occur only at specified junctions;
- the necessity of obtaining an estimated travel time for each location of the study area; and
- the large extent of the travel-time surface to be developed, as in developing countries the major accessibility issue relies on linking remote rural areas to major urban centers.

This chapter addresses these requirements by providing a simple, flexible, and reliable methodology that can be applied to other regions facing similar travel-time modeling issues. Our working assumption is that an integrated approach, solving a multimodal transportation network along with developing a travel-time surface for the entire study area by taking into account off-road travel, can outperform standard methodologies to estimate the actual travel time in developing countries.

The chapter is structured as follows. Next sections present the study area, followed by the description of the proposed hybrid modeling and the validation. After the results discussion, modeling outputs demonstrate relevant modeling applications.

3.2 Study area

The travel-time modeling was implemented in the north-east of the DRC (over an area of approximately 12,200 km^2), within the *Province Orientale*, the biggest DRC Province covering slightly over 400,000 km^2. According to the Gridded Population of the World Version 3 (GPW-v3) (Balk and Yetman, 2004), the human population in the study area is about 751,400 inhabitants. This specific region around the city of Kisangani is benefiting from a growing interest in understanding the impact of major cities' accessibility from surrounding human settlements (Bamba *et al.*, 2010).

Kisangani, considered the destination point in this study's travel-time modeling, is the main city in the study area and is located along the Congo River. The *Référentiel Geographique Commun* (RGC) database has inventoried approximately 200 villages located along the main road network (Blaes, 2009); while from GPW-v3, 35% of the study area total population was found at more than 5 km buffer from all network segments.

The motorcycle is the most common means of transport on the road. The road network density, however, is very low (on an average 0.03 km per km^2). The road network is sub-divided into national and local roads denoted as NR and LR, respectively (Figure 3.1).

Approximately 30 villages are located along the section NR4 that connects Kisangani to Banalia. These villages are known for supplying Kisangani with

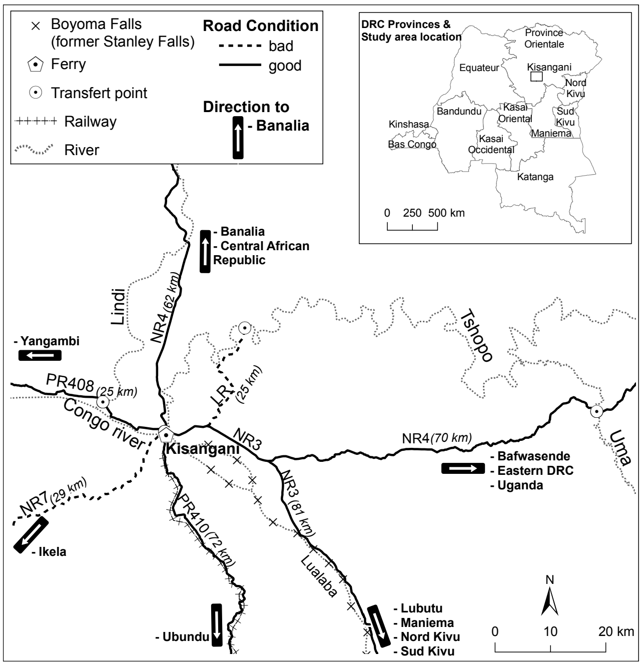


Figure 3.1: Study area transportation network.

agricultural products while they benefit from manufactured products. The other NR4 section connects Kisangani to Bafwasende and constitutes one of the major West–East transportation axes at the DRC national scale. Other villages are located along the NR7 (±12 villages), the prioritary road (PR) 410 (±25 villages), the LR (±15 villages), and the PR408 (±25 villages).

Certain navigable rivers complement the study area road network. Among these are a portion of the Congo River (23 km), which flows approximately 1,750 km from Kisangani downstream to Kinshasa, a portion of the Lualaba River (72 km), which becomes the Congo River downstream from Kisangani after Boyoma Falls, a portion of the Tshopo River (224 km), which joins the eastern regions of the study area to Kisangani, and a portion of the Uma River (67 km).

Other components of the transportation network consist of one ferry, three transfer points, and one bridge on the Tshopo River along the NR4. The ferry, which has a port located in Kisangani, allows for crossing the Congo River. Transfer points between the road and the river network are located at the following junctions: PR408 at the Lindi River, LR at the Tshopo River, and NR4 at the Uma River.

3.3 Methodology

3.3.1 Overview

The travel-time model developed in this chapter combined the resolution of a multimodal network with *off-network* travel to produce a travel-time surface that depicted travel-time values from any location in the study area to selected destinations. The multimodal network addressed network specificities such as the travel speed by network segments and possible travel restrictions or barriers. Walking-based *off-network* travel took into account land friction and was optimized to join the transportation network at the most suitable transfer location.

The hybrid travel-time model was based on three main points (Figure 3.2a). The first used the multimodal transportation network for extracting route lines. The second point emphasized the iterative process for computing the travel time for areas located farther away from the transportation network segments. For these areas, we computed the travel time to the nearest destinations by combining both *off-network* and network travels. In the literature Brainard *et al.* (1997); O'Sullivan *et al.* (2000); Liu and Zhu (2004a), these two components were computed separately, while in this research, they were jointly optimized. The implemented decision rule chose the most suitable point for joining the transportation network from *off-network* travel in order to reach the nearest destination. The last point included the introduction of a friction surface into the modeling process and the use of regular hexagons such as primary units of analysis.

Assuming that d is the side length of a square cell, cardinal movements between centroids of adjacent cells correspond to a distance of d. The corresponding distance for diagonal movements is in turn computed as $d \times \sqrt{2}$. Meanwhile, the distance between centroids of adjacent hexagons is identical in all possible directions. This characteristic of the hexagon configuration (i.e., same distance between centroids of adjacent hexagons) was advantageous for modeling purposes (compared to a square grid), as demonstrated by de Jong and van Eck (1996); de Sousa *et al.* (2006); Birch *et al.* (2007). Using a grid of hexagonal cells simplified the algorithm scripting. However, hexagons by nature are limited to only six neighbors compared to square cells that have eight neighbors.

In this study, the area of the hexagons used was about 866 m^2. The distance between the middle points of the opposite edges was 1,000 meters. The hexagonal tessellation of the study area was stored as a native ESRI shapefile.

3.3.2 Generation of the travel-time surface

Transportation-network-specific rules were used for solving the multimodal network. Route lines that measured the minimum travel time to and from selected destinations were extracted from the multimodal network and integrated into the second modeling step. Using linear referencing techniques (Curtin *et al.*,

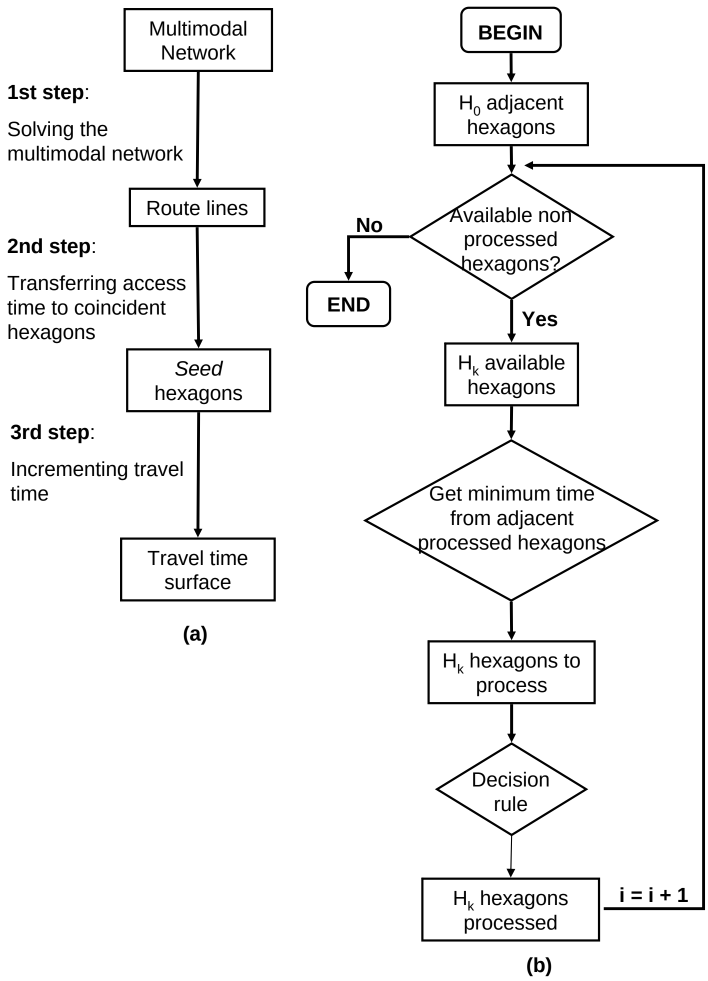


Figure 3.2: (a) Three main modeling steps, and (b) detailed steps of the iterative process for computing off-network travel time.

2007), travel-time measures from route lines were transferred to coincident hexagon objects. The set of hexagons (H_0 hexagons) resulting from this operation were considered *seed* objects for the further iterations aimed at computing the travel-time values for areas located farther away from network segments.

Travel-time values for areas lying within network segments were computed by iteratively incrementing seed hexagons' time values with a typical walking time ($INCR$) for crossing one unit of analysis. The increment value ($INCR$) (in hours) for moving from the set of H_0 hexagons to their neighbors was given by the following expression, adapted from van Wagtendonk and James (1980):

$$INCR = \frac{d}{v_0 \times e^{-ks}} \tag{3.1}$$

where d is the distance between centroids of adjacent hexagons (in meters), v_0 is the pedestrian walking speed over flat terrain according to the land-cover type (in meters per hour), s is the slope (in meters by meters), and k is a factor defining the effect of slope on walking speed.

Land cover types used in this research are described in Vancutsem *et al.* (2008). Pedestrian walking speeds by land-cover type, adapted from Nelson (2008), are summarized in Table 3.2. The slope was derived from SRTM V4 data. For the sake of simplicity, and given the relative flatness of topography in the study area (the maximum slope being slightly less than 10%), we opted for using a mean value of k (k = 3), which corresponded to the slope effect on downhill and uphill movements. Other authors also considered admissible the use of this mean value (Watts *et al.*, 2003; Nelson, 2008).

Table 3.2: Travel speed by land-cover types adapted from Nelson (2008).

Classes	Speed (meters per hour)	Classes	Speed (meters per hour)
Edaphic forest	1,000	Aquatic grassland	1,000
Dense moist forest	1,000	Old secondary forest	1,000
Young secondary forest	1,667	Rural complex	2,500
Forest savanna mosaic	1,250	Steppic savanna croplands	1,667
Shrubland	1,667	Grassland	1,667
Miombo	1,000	Agriculture	1,000
Savanna woodland	1,667	Tree savanna	1,667
Swamp grassland	1,000	Submontane forest	1,000

The model assumed that movements through the tessellation of hexagons between successive iterations connected centroids of processed hexagons to centroids of adjacent non-processed hexagons, leading to *zigzag* patterns as depicted in Figure 3.3.

One run of the iterative process corresponded to a series of operations (Figure 3.2b) summarized hereafter.

The travel time for H_0 hexagons crossed by network segments was obtained from route lines. Other hexagons covering the study area were assigned an actual travel time (ATT) equivalent to ∞. The hexagons also stored information related to the restriction for traversal during the iterative process (i.e., RH). If $RH = 1$, the actual hexagon could not be connected to *off-road* travel during the iterative process. This was the case for H_0 hexagons located along the river network that could be connected to *off-road* travel at existing transfer points only.

Second, the actual travel time of all hexagons directly adjacent to H_0 hexagons (i.e., H_i hexagons, with i corresponding to the current iteration) was updated according to the following decision rule:

if RH_i ***and*** $RH_{i-1} \neq 1$ **then**
  if $ATTH_{i-1} + INCR < ATTH_i$ **then**
    $ATTH_i = ATTH_{i-1} + INCR$
  else
    $ATTH_{i-1} = ATTH_i$
  end
end

Prior to the following run of the iterative process, H_i hexagons became new *seed* hexagons. The following run was then executed using the same procedure until all hexagons were processed.

The result of this iterative algorithm was a set of hexagons that stored the minimum time for an optimized travel to reach the nearest destination, from any location of the study area.

3.3.3 Model parameters calibration

Travel speeds by motorcycle were calibrated from field-based surveys. More than 150 villages along the road network and approximately 500 km of road tracks were surveyed from December 2010 to January 2011 using Garmin 60 CSX GPS (accuracy < 10 meters) . Travel times for reaching Kisangani by motorcycle from the 150 villages were also recorded. As such, travel speeds by road segment were derived from univariate analysis (with the intercept set to 0) by randomly selecting 75% of the geo-referenced locations for each road

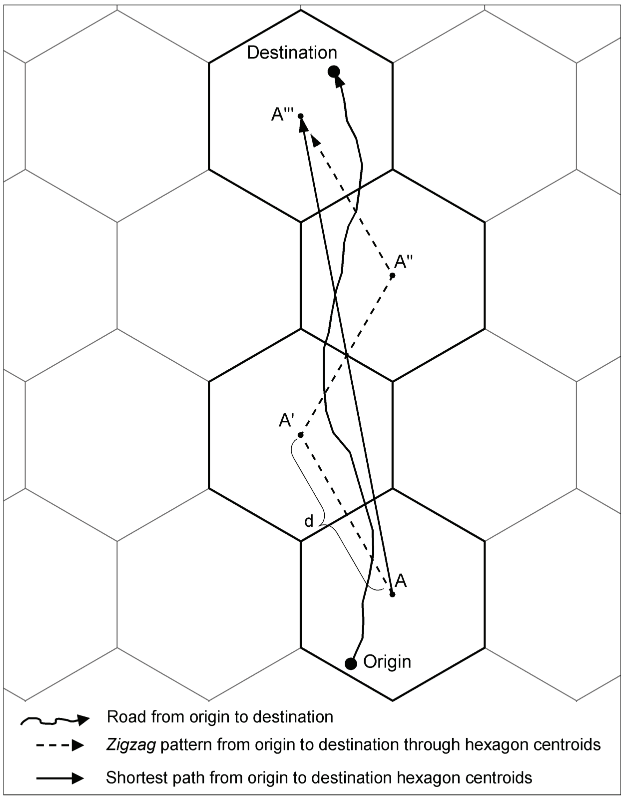


Figure 3.3: Path approximation from an origin to a destination through the hexagons tessellation.

axis. Derived motorcycle travel speed per road segments showed small variations that ranged from 30 to 36 km per hour (Table 3.3).

Table 3.3: Travel speed estimations for selected road segments in the study area.

From Kisangani to	Selected villages (n)	Speed (km/hour)	P value	F-stat.	R^2
Banalia (NR4)	32	35.971		4,801	0.994
Opala (NR7)	23	30.408	$< 2.2e\text{-}16$	4,604	0.995
Ubundu (PR410)	26	35.187		9,837	0.997
Yangambi (PR408)	24	30.471		3,084	0.993

In the absence of precise field-based data corresponding to travel speeds on rivers for non-motorized pirogues, speeds were fixed to 15 km per hour for downstream and 10 km per hour for upstream travels. The waiting time for road-river interchanges was fixed to 1 hour, and the ferry speed was set to 15 km per hour.

3.3.4 Validation data set

Villages not included in the travel speed by road axes calibration were used for comparing field-based and modeled travel time to Kisangani. These 34 villages were located up to 80 km from Kisangani.

Cross-comparisons between outputs from a nearest neighbor-based modeling and the hybrid travel-time model were also conducted, for the three emblematic situations stressed in Section 3.2. These real-world selected cases illustrated that (1) the shortest path is not always the fastest one due to transfers between network components, which are restricted to specified junctions (Figure 3.4a — Case 1), (2) the closest network is not always the fastest route to the destination (Figure 3.4a — Case 2), and (3) the nearest neighbor does not always offer the fastest route due to land friction (Figure 3.4b — Case 3).

Nearest-neighbor travel-time modeling for the three cases was implemented using the same parameter values (i.e., travel speed by road type, river type, *off-network* type, and incorporating the friction surface).

3.4 Results

The transportation network described in Section 3.2 constitutes a multimodal network with specific rules as follows:

- navigation on the Lualaba River is no longer possible between Kisangani and Lubutu

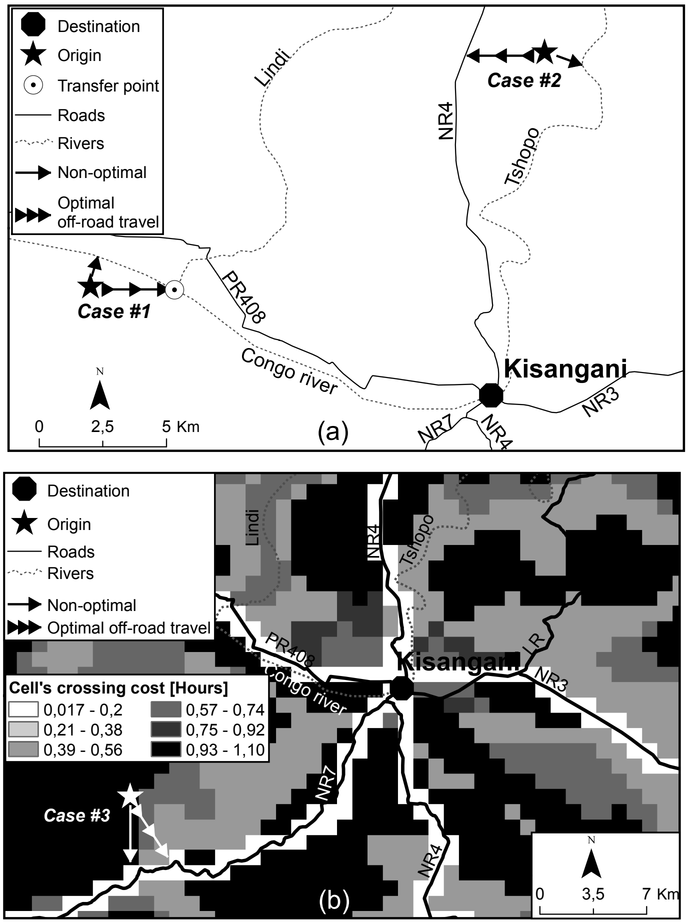


Figure 3.4: Optimal routes according to the hybrid model and suboptimal routes according to nearest network-based modeling for cases 1 and 2 (a) and case 3 (b). Cell's crossing cost values are derived from land-cover classes and the slope.

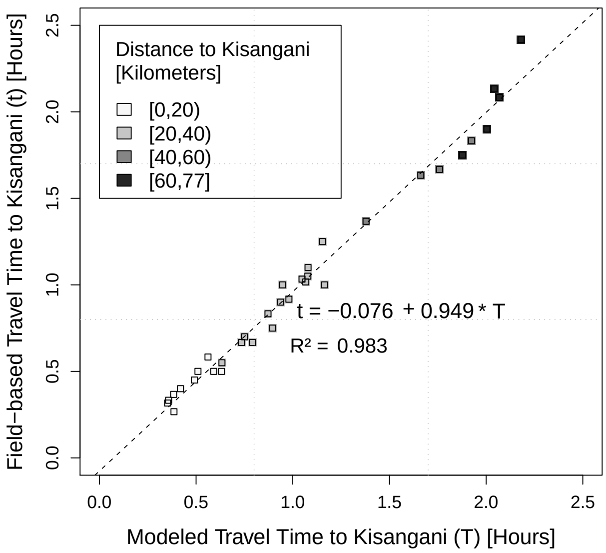


Figure 3.5: A scatter plot showing modeled *vs.* field-based travel time to Kisangani. The dashed line represents the regression line and each dot represents a village location.

- *off-road*—river interchanges are only possible at transfer locations
- there is no transfer between the Tshopo River and the NR4 at the bridge location
- travel speed on the river network varies according to upstream and downstream travel.

The hybrid travel-time model was applied to the study area by taking into consideration specific rules stated here above. The whole study area was covered after around 2 hours of computation.

Validation results based on the comparison between measured and modeled travel time from the 34 villages to Kisangani are illustrated in Figure 3.5. Values obtained by simulated travel times were in close agreement with field-based surveys ($R^2 = 0.98$; $F_0.95(1, 32) = 1{,}902$). The slope confidence interval of the regression line at $\alpha = 0.05$ was close to one, i.e., [0.988 - 1.085]. Corresponding values for the intercept, i.e., [-0.134 - -0.018], indicated that the travel-time model slightly overestimated observed values. This could be attributed to hexagons discretization and the *zigzag* patterns from moving between hexagons' centroids (Figure 3.3).

Results from the comparison between nearest neighbor-based modeling and the hybrid travel-time model are summarized in Table 3.4.

Table 3.4: Comparisons of results from nearest-network modeling (NNM) and the hybrid model (HM) for selected real-world cases.

Case	Distance to Kisangani (km)	NNM (hours)	HM (hours)	Absolute difference (hours)
Case 1	22	1.73	2.33	0.60
Case 2	10	1.07	0.70	0.37
Case 3	24	4.90	3.99	0.91

Case 1 shows that joining the Congo River from *off-road* travel is possible only at a transfer point. Therefore, the optimal *off-road* travel (and accordingly the time for reaching Kisangani) is longer. Case 2 shows that it may be optimal to take a longer *off-network* route in order to reach a faster network segment, instead of reaching the nearest network segment such as previously modeled. In fact, the travel speed on NR4 is 130% faster than the travel speed on the Tshopo River. As a consequence, computing the travel time to Kisangani according to a nearest network rationale can lead to erroneous results. This is particularly the case for origin locations situated farthest away from Kisangani and located closer to the Tshopo River than to NR4. The last case (Case 3) illustrated in Figure 3.4b shows that the shortest *off-network* travel is not always the most optimal due to land frictions of certain land-cover types or topography. Therefore, the use of a nearest network-based modeling in this latter case could introduce a severe bias in the estimated travel time to the destination. In turn, the developed algorithm optimizes both network travel and *off-network* travel, avoiding higher surface friction while also minimizing the travel time to the destination.

We also compared the results of the generated travel-time surfaces for a subset of the study area covering approximately 3,230 km^2. From this subset, the optimal itinerary to Kisangani reveals a *rerouting* point located on the Tshopo River (Figure 3.6b) that optimizes the travel time to Kisangani using a transfer point located at the junction of the NR4 and the Uma River. As a consequence, the transfer time at the junction Uma-NR4 will drastically impact the resulting travel-time surface in this region.

The results of cost-distance-based modeling (Figure 3.6c), however, do not highlight the expected advantage of both the rerouting and the transfer points. As opposed to the developed hybrid model, the cost-distance-based model does not clearly show how coming from Kisangani along the Tshopo River, the travel time increases till a maximum at the *rerouting* point. The cost-distance-based model neither shows how after the *rerouting* point, the travel time starts a declining phase along the optimal route for reaching Kisangani along both the Uma River and the NR4, passing by the transfer point.

Also, cost-distance-based modeling does not distinctly illustrate the transfer point impact on the resulting travel-time surface. Indeed, the pattern of the

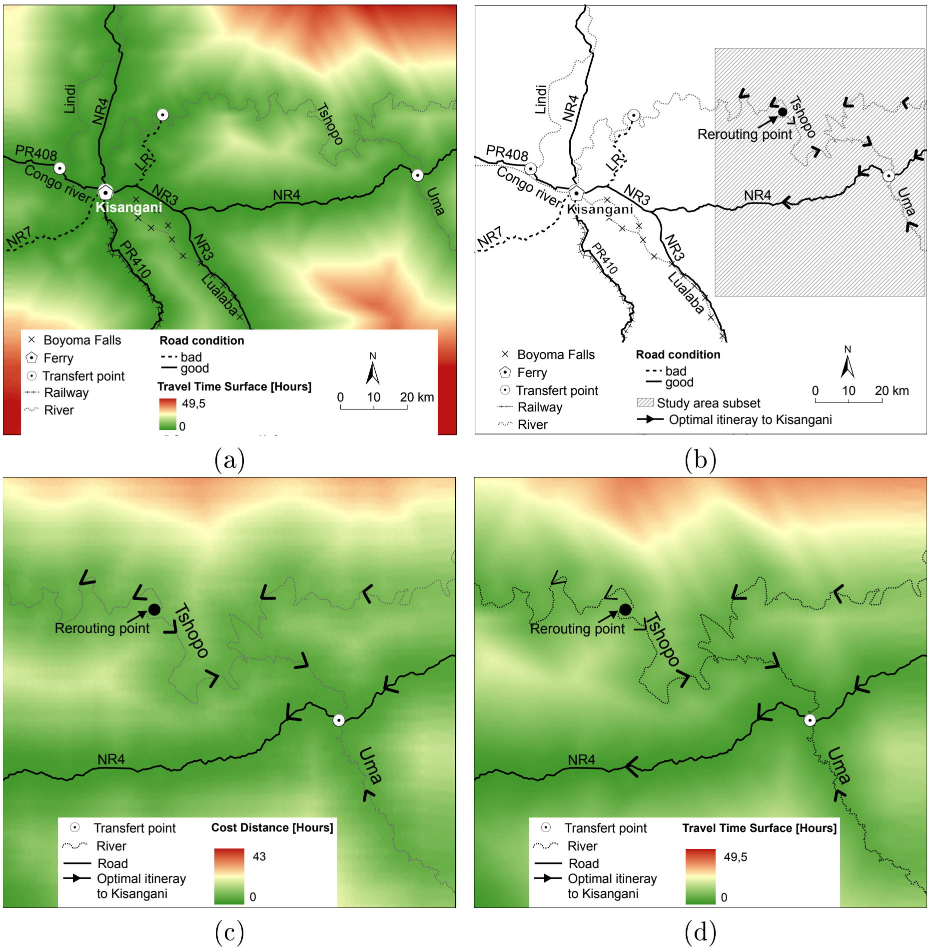


Figure 3.6: (a) Travel time to Kisangani for the entire study area: in this simulation, *off-road* travel is allowed to join the road network at any location, while its junction with the river network is only feasible at transfer points. (b) Optimal itinerary to Kisangani for network segments from origins located within a study area subset. (c) and (d) are respectively the cost-distance-based and the hybrid model-based travel time to Kisangani for the subset outlined in 6b.

corresponding surface is most similar to corridors along the transportation network. In turn, the travel-time surface pattern from the developed hybrid model (Figure 3.6d) clearly shows the impact of the transfer point. At the same time, areas located farthest away from both the transportation network and Kisangani result in higher travel times compared to areas located both close to the transportation network and farthest away from Kisangani. Clearly, *off-network* travel, along with land frictions, is optimized for joining the transportation network at the best locations, minimizing the travel time to the destination.

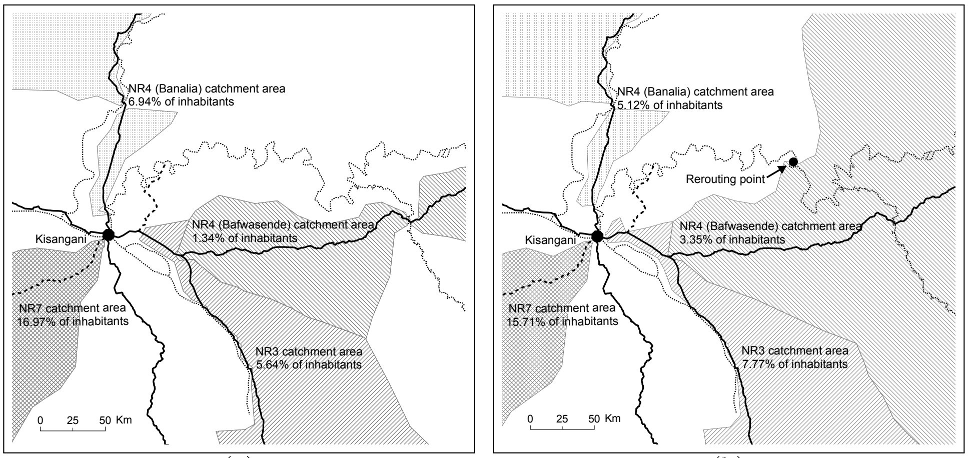
3.5 Discussion

The presented methodology is a hybrid approach for estimating the travel time from any location of the study area to selected destinations over large regions. This approach is particularly useful for, but is not restricted to, regions located within developing countries where *off-road* travel is particularly important. The developed hybrid model was applied in the north-east of the DRC, over an area covering approximately 12,200 km^2.

The use of hexagons as primary units of analysis brought some valuable simplifications to the iterative algorithm implementation. The algorithm developed here could also be adapted to square grid cells and resulting travel times and processing time should be compared to those obtained using the hexagonal tessellation. Such a comparison would identify the best trade-off between reducing the solution space discretization and the processing time for larger scale or finer resolution operational implementation.

The final output of the model is stored as a native ESRI shapefile that has open specification for data interoperability. The modeling output could be converted into a raster format for representation purposes or further analyses, but this could degrade the original information (Congalton, 1997). Yet, the rasterization of the model output is not mandatory because further GIS analysis and/or result display could be carried out using the original vector format. The presented method finds the optimal route to reach the nearest destination using the assumption that connections between specific travel modes (i.e., *off-road*—river and river-road connections) are restricted, occurring only at transfer points. Most studies, particularly those relying on a nearest network-based modeling procedure, simplify this reality. However, other determinants of connections between travel modes, such as trip distances, individual behavior, means of transport availability, transport and/or opportunity costs, and population income (Cervero and Radisch, 1996; Srinivasan and Rogers, 2005; Scheiner and Holz-Rau, 2007; Scheiner, 2010) have not been considered in this research.

The method developed here also determines a travel-time estimation for areas around transfer locations. The resulting travel-time surface clearly shows that transfer locations become somehow *secondary destinations* before reaching the targeted destination. The impacts of transfer locations on travel time



(a) (b)

Figure 3.7: Catchment areas for selected roads in the study area from: (a) cost-distance-based and, (b) hybrid model-based travel time to Kisangani.

are especially prominent at junctions located between transportation segments that have different travel speeds. However, transfer locations do not exhibit their peculiar function in a nearest network-based modeling framework since multimodal characteristics are not taken into account.

As a result of this more realistic travel-time modeling approach, it is possible to derive the catchment area for each transportation network segment. Such a catchment area can provide valuable information for decision-makers. For instance, it can be used to derive the number of people who will benefit from a given road improvement (Figure 3.7). It is worth noting that a catchment area computed from a cost-distance-based model can drastically differ from a catchment area derived from the presented hybrid approach. This is mainly due to the fact that a cost-distance-based computation is not primarily equipped for handling multimodal travel-time modeling (DePencier, 2008). For example, by comparing Figure 3.7a and b, the NR4 (Bafwasende) catchment area from a cost-distance-based model is clearly underestimated. In fact, the cost-distance-based catchment area does not take into account the contribution of network segments from both the Tshopo and the Uma rivers — due to the presence of a *rerouting point* — that have to utilize NR4 (Figure 3.6b) in order to minimize the travel time to Kisangani.

The presented hybrid model also shows how the travel time to a given study site might vary according to both different time intervals and travel modes. Figure 3.8 illustrates the travel time to a secondary town for pedestrian, motorcycle, and combined motorcycle and pirogue transportations. The different maps show the large differences between areas that can access a town within 2 hours according to the various transportation modes. As shown in Figure 3.8f, the highest differences between transportation times occur at lower time intervals, for example, in 30 minutes and 1 hour isochrones. This results from travel speeds differences on the network. It is interesting to note, however, that at higher time intervals, these differences are flattened. This is due to a greater contribution of *off-network* travel, which is walking-based for all travel modes. As a consequence, shifting from one travel mode to another will mainly impact the travel time for areas located closest to the transportation network. In turn, travel times for areas located away from the transportation network will mainly rely on the *off-network* travel easiness. This supports the combination of both the network-based travel time and the friction surface component such as achieved in the developed hybrid model.

By looking at the Figure 3.8d in particular, some inclusions in the resulting travel-time surface are visible, for example, along the river at the south of Wanie Rukula. These inclusions illustrate the realism of a combined approach that optimizes jointly network-based travel time and a friction surface at the same time; areas close to the transportation network have not always a lower travel time to the destination. These inclusions, of course, are not modeled by cost-distance algorithm.

The travel-time model was implemented for the whole DRC, using 165 main towns such as destinations. These were selected based on their total population in 2005 (i.e. $\geq$ 50,000 inhabitants). Road condition was used to determine the travel speed: 60 km/h for roads in good condition, 40km/h for roads in medium condition, and 30 km/h for roads in bad condition. Ferries, ports, railroads, and all navigable rivers were also combined to obtain a fully-modeled and topologically-validated multi-modal transportation network. A detailed description of the national transportation network used is provided in Section 4.2.3. The nationwide scale travel-time map is presented in Figure 3.9, which indicates that any location within DRC can be reached from the nearest town within a maximum of 3 days, according to the model assumptions. Such information are valuable for land-use planing, services access and resources management.

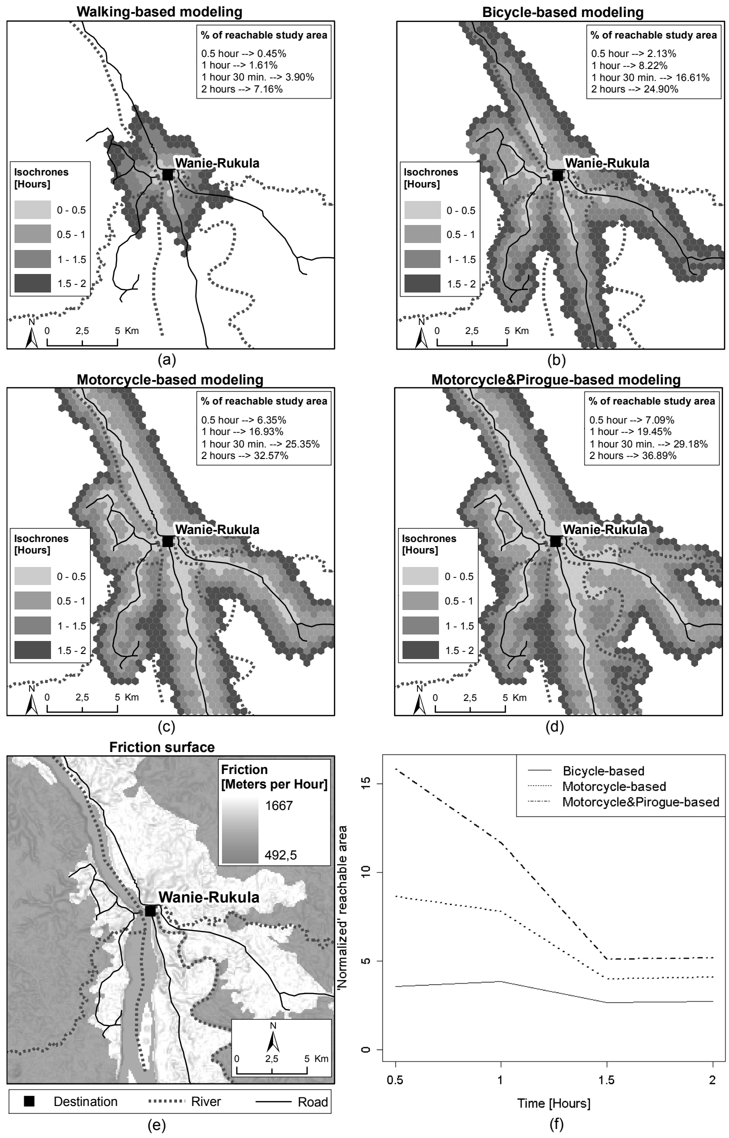


Figure 3.8: (a) to (d): Reachable areas from Wanie Rukula using different travel modes. Travel speeds are: 3 km/h (walking-based), 10 km/h (bicycle-based), 15-20 km/h (motorcycle-based, according to roads condition), and 15 km/h (pirogue-based). In the above simulations, it was assumed that *off-network* travel can join the transportation network at any location on the network, and not only at specified stops. (e) Friction surface values. (f) Reachable areas from bicycle-based, motorcycle-based, and motorcycle&pirogue-based modeling normalized by walking-based modeling.

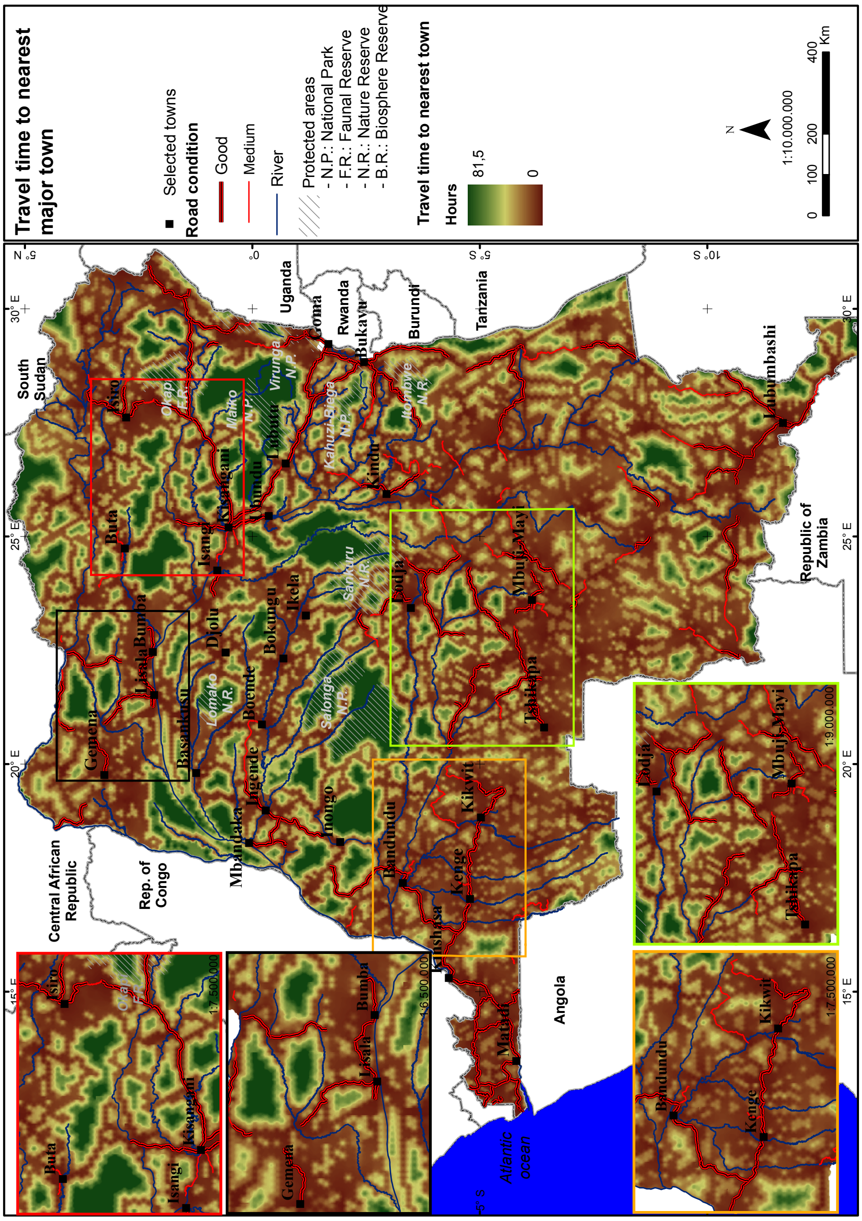


Figure 3.9: Travel time to nearest major town or big village for whole DRC.

3.6 Conclusion

The developed hybrid modeling approach successfully combined a fully modeled multimodal network and generated a travel-time surface for the entire study area. The hybrid modeling performs better than the existing vector-based methods that do not account for *off-network* travel, especially for regions that suffer from a lack of a formal transportation network as is the case in many developing countries. The hybrid model also provided precise estimates at a local scale, which is paramount because daily access to market, school, or hospital in these countries is mainly walking-based and therefore highly sensitive to travel-time variations. While the hybrid method supports multimodal aspects, it remains robust also for regions where only the road network is used for transportation. Also, because it is possible to generate a travel-time surface 'to' the destination as well as a travel-time surface 'from' the destination, the hybrid model adds value for exploring, in terms of accessibility, the relationships between origin and destination locations. Meanwhile, the travel-time modeling developed here can also be used to derive certain accessibility measures described by Geurs and Van Wee (2004), such as utility-based accessibility measures.

However, the developed model did not take into account the seasonal variability of roads' condition, which can drastically differ between dry and rain seasons. Although there is no precise information about which road segment may become impassable during the rainy season, such event can be simulated using topography information from SRTM data (Gleyze and Rousseaux, 2003). Model shortcomings such as the waiting time for shifting from one travel mode to another has to be investigating since a fixed waiting time is unlikely to correspond to all possible circumstances at the national level. In addition, the model assumed that the nearest town was the most suitable without taking into account the travel purpose; which can allow ranking all possible destinations to focus only on travels providing the desired outcome.

Overall, the model developed in this Chapter can also support decision-making process for the development of secondary towns, or also to objectify planning strategies such as a road rehabilitation (for making more accessible remote rural areas), which is crucial in most of the developing countries. Impact assessment of infrastructure development can therefore be anticipated to investigate various scenarios of investments. Quantifying the travel time in a realistic way for the DRC paves the way for more objective information, supporting strategic land-use planning and regional development.

Chapter 4

Drivers-based Forest Cover Change Analysis and Modeling

Abstract

Since the 1990's, the DRC is facing increasing rates of deforestation and forest degradation. A major part of the DRC human population depends almost exclusively on agriculture and forest products for its livelihood. The main driving forces of the forest cover loss are population growth, the expansion of agricultural areas, the proximity to major towns and big villages, fuelwood collection, and charcoal production. Modeling approaches that attempted to simulate the forest cover changes in DRC are, however, hindered by the scarcity of the data which can suitably describe the underlying driving forces of change. The model developed in this chapter uses a set of spatial variables (i.e., the human population distribution, the accessibility to major towns and big villages, the proximity to already existing secondary forests, and the distance to protected areas) for simulating the forest cover changes at the DRC nationwide scale from 2005 to 2035. 70% of changes between 2000 and 2005 were located in a daily journey from and to the nearest major town and the BAU scenario indicated that the DRC population will at least double in 2035, leading to an annual rate of forest cover loss of about 0.48% between 2005 and 2035. The forest conservation scenario indicated that the DRC will likely remain a high forested and low deforested country in 2035 if mitigation actions to reduce the forest cover losses are undertaken. Impact assessment of road construction and rehabilitation as well as a strategic land-use planning at the national scale are therefore required so as to allow conciliating the economic development with a sustainable use of the forest resources.

4.1 Introduction

A major part of the human population living in tropical sub-saharan countries depends almost exclusively on agriculture and forest products for its livelihoods. Therefore, the forest change dynamics in these regions mainly result from the basic socioeconomic needs, such as the croplands and the domestic energy needs (Zhang *et al.*, 2006; Kamusoko *et al.*, 2009; Tollens, 2010; Nackoney *et al.*, 2013).

Defourny *et al.* (2011) found that agricultural expansion, fuelwood extraction, population density, and the proximity to major towns and big villages were positively correlated with the annual rates of deforestation and forest degradation in the DRC. These authors also concluded that protected areas played a strong role in limiting the forest cover loss.

Modeling approaches that attempt to simulate these forest cover changes are, however, hindered by the scarcity of the data which can suitably describe the underlying driving forces of change (Zhang *et al.*, 2002; Tollens, 2010). Indeed, the proximity to roads and major cities is most often computed from distance-based models which do not adequately take into account alternative transportation systems (Nelson, 2008), while population distribution data commonly used have shown a poor performance in developing countries (Chapter 2). In addition, the drivers of forest cover change most often interact in a complex way that can hamper a proper modeling of the change processes (Mas *et al.*, 2004).

Most important, the variables used in forest cover change models have to accurately translate the drivers of change. However, the various operations on these variables — ranking (Zhang *et al.*, 2006), standardization (Sangermano *et al.*, 2010), aggregation (Pontius Jr *et al.*, 2001) — such as implemented in many models can somehow conceal the meaning of the drivers. The possible correlation between these variables has also to be analyzed since the most commonly used such as the distance to roads, human settlements, and rivers (Mertens and Lambin, 2000; Mas *et al.*, 2004; Laporte *et al.*, 2007; Khoi and Murayama, 2011) are sometimes highly correlated (Mas *et al.*, 2004), which can lead to model over-fitting.

Models of forest cover loss based on Markov chain typically involve the analysis of two initial land-cover maps that can be referred such as the t_0 and the t_1 reference maps. These commonly share the same legend and spatial resolution. They are afterward compared at the pixel level in order to differentiate the pixels that belong to the same land-cover class (i.e., land-cover persistence) from those enclosed into different categories (i.e., land-cover transitions or changes) (Eastman *et al.*, 2005). Land-cover transitions, in particular, inform on both the quantity and the location of changes that serve as the basis for simulating the future forest cover changes.

Spatially-explicit land-cover change models are broadly made of three components that are: (1) the amount of change, (2) the potential for change, which evaluates the likelihood of each location in the study area to change from one land-cover type to another, and (3) the spatial allocation, which determines the locations where the amount of change will be allocated.

4.1.1 Amount change

The amount of change can be derived from the land-cover changes that occurred between the t_0 and t_1 reference maps, by relying on Markovian or semi-Markovian transition matrices (Brown *et al.*, 2000). These are computed by cross-tabulating the two maps in order to retrieve areas that transitioned from one land-cover type to another. Markovian transition matrices therefore provide both the total amount of change between reference maps and the amount of change for each transition that will be used to simulate the future changes. Given that Markovian transition matrices assume that the amount of change do not vary over time, an alternative is the quantification of the future land-cover changes from economic or socio-economic models, which can include regional and global scale driving variables of land change (Schneider *et al.*, 2011). These models are generally based on the von Thünen's land-rent theory that states that each parcel of land is converted into the type of use yielding higher profits (Schneider, 1995; Chomitz and Gray, 1996). Soares-Filho *et al.* (2002) used a different approach to compute the amount of change by allowing future land-cover transition rates to vary over simulations. In doing so, the allowed amount of change per transition was constrained by a so-called saturation value, which reflected that land-cover transitions were function of political, social, economic, and natural circumstances (Verburg *et al.*, 2004a). As subsistence activities are the main drivers of forest cover change in DRC, Markovian transition matrices were selected for this study.

4.1.2 Potential for change

The likelihood of a specific location to change from one land-cover type to another is mainly dictated by factors such as its suitability for the transition in question, and the presence of driving forces of change (Eastman *et al.*, 2005). For each considered transition, this potential is generally obtained from the analysis of changed pixels between t_0 and the t_1 reference maps with regards to the driving forces of change. These can indeed deliver substantial information for predicting the future locations of land-cover transitions since they highlight the different landscape dynamics that occurred in the past (Mertens and Lambin, 2000; Echeverria *et al.*, 2006; Eckert *et al.*, 2011).

Two types of methods are generally used to compute the potential for change. They vary according to the integration of predictors of change into the modeling process. The first type relates to the assessment of each variable separately, followed by an aggregation procedure in order to yield a single statement of the potential for change (White and Engelen, 2000; Pontius Jr *et al.*, 2001). The second type uses the all set of variables at once for evaluating their combined effect on the observed changes. Logistic regressions and artificial intelligence methods are among the most popular techniques of this latter type (Berry *et al.*, 1996; Eastman, 1999; Verburg *et al.*, 1999; Pijanowski *et al.*, 2002).

Logistic regressions can both link the presence and the absence of land-cover data with any form of explanatory variables. They yield a probability statement that can be interpreted such as a potential for land-cover change (Gil Pontius Jr. *et al.*, 2001; Verburg *et al.*, 2002). Typically, the occurrence of a given land-cover type is regressed against a set of predictors, and the results from the regression analysis allow constructing the potential for change for each land-cover class. The simulation of future land-cover changes is achieved by updating the potential for change according to the values of the predictors of change that are expected to vary over time.

Artificial intelligence (AI) methods, for their part, seek to develop a framework that attempt to mimic the human intelligence without claiming an understanding of the underlying change processes. Artificial neural networks (ANN) are part of the AI methods that are used to compute the potential for change. They can potentially handle the complex relationships existing between the predictors and the corresponding changes (Mas *et al.*, 2004; Eastman *et al.*, 2005). The most common ANN learning method is the back propagation of the errors, which uses a set of areas of known changes (i.e., the training data set) to determine the errors by comparison with the model output (Pijanowski *et al.*, 2002; Khoi and Murayama, 2011). These errors are propagated through the network in order to iteratively adjust the weights assigned to each variable of change.

4.1.3 Spatial allocation of change

The spatial allocation determines the locations where the amount of change will be allocated. Generally, the selected locations are those which exhibit the highest potential for change values (i.e., typically above a specified threshold) according to the modeled transition. The total area of selected locations (e.g., pixels) corresponds to the amount of land-cover change.

An alternative to the deterministic specification of a single threshold is the iterative thresholding. In this case, a predefined potential for change value is set for initializing the iterations. This value is then adjusted over model simulations until the required quantity of change is met. Alternatively, potential for change values can be reclassified into percentile ranges that serve to fix the threshold for change (Pijanowski *et al.*, 2002). Change allocation can also spatially constrain the locations of change so that they lie within fixed boundaries that can be buffer zones around specific land-cover classes (Pontius Jr *et al.*, 2001).

Neighborhood interactions that can be inherent to land-cover changes (Verburg *et al.*, 2004a) are not explicitly taken into account when the change allocation is based solely on the potential for change values. Models that explicitly account for these interactions commonly implement cellular automata (CA) simulations. These allow to model the transition of a cell from one land-cover type to another depending on its actual land-cover type and the land-cover types of its surrounding cells. The major effort consists therefore into matching the transition rules of CA models with the reality (Torrens and O'Sullivan, 2000; Verburg *et al.*, 2004a)

Overall, the research area on simulating the land-cover dynamics faces three major interrelated challenges. These refer to the three broad components of land-cover change models. Indeed, a good approximation of the potential for change makes the allocation process quite simple since the potential locations of change are already well determined. Also, if the allocation procedure is well conducted, then estimating the potential for change can be simplified. Similarly, if the amount of change is estimated in a spatially-explicit manner, then the process of computing potential for change maps can be alleviated.

The first challenge then relies on estimating the total amount of change in the future. While the corresponding estimates derived from Markovian transition matrices can be seen as acceptable on a short term period, long term predictions have to account for internal and external socio-economic driving variables. The second challenge refers to the evaluation of these predictors of change, since they are used afterward to derive the potential for change. While the variables have to meaningfully reflect the proximate causes of change, they can also be difficult to retrieve in the poor data context of many sub-Saharan countries. Finally, determining the locations where the amount of change will be allocated is critical since it can be contingent on the method used, e.g., selection of highest potential for change values or stochastic selection within a range of potential for change values. Validating the simulated map of change for a t_2 period with the corresponding t_2 land-cover map is an objective assessment of modeling outputs.

This chapter develops a spatially-explicit and dynamic model that simulates the future forest cover losses in DRC from 2005 to 2035. The year 2035 was chosen since it is the ending year of the current DRC national REDD+ strategy[1]. Simulations are yearly-based and are conducted for two selected scenarios that are detailed in Section 4.3.1. They differ each other from the amount of change estimation. These scenarios were chosen given the few existing data for the DRC to efficiently achieve spatially-explicit simulations of forest cover changes.

Meanwhile, other possible scenarios simulating the development of agribusiness or major macro-economic changes could be used, but these were discarded since they would turned out to be extremely speculative in the DRC context due to the very few available information (Megevand *et al.*, 2013). The first scenario is a forest conservation model based on the assumption that the total amount of future land-cover changes will be of same magnitude such as observed during the 2000-2005 period. Such scenario would suggest implement-

[1] `redd.cd/fr/redd-en-rdc/mission` - [Accessed 20 March 2013]

ing land-use planning policies possibly supported by REDD+ like initiatives. The second scenario is a business-as-usual (BAU) scenario that only accounts for the population growth. Outputs from these scenarios are expected to be useful baseline information for building the required national reference level of emissions in the REDD+ framework (Karsenty and Ongolo, 2012; Lu and Liu, 2013).

The data sets used (Section 4.2) to simulate the future forest cover losses according to both scenarios were selected based on the model requirements to (1) compute the amount of change, (2) estimate the potential for change, and (3) achieve the spatial allocation of change.

4.2 Data sources

The data sets used for the developed model are described here below.

4.2.1 Land-cover maps

The FACET maps were used in this study since they are to-date the only available data sets providing three land-cover maps for three distinct time-periods for the whole country (Figure 4.1 - *Step I*). While these maps have their own errors that were discussed in the Chapter 1, they are based on the same legend and were obtained from a single methodology. They are the only wall-to-wall maps that can be cross-compared.

The FACET data set quantified the forest cover and its losses in the DRC for the decade 2000-2010 at 60-meter spatial resolution. This was achieved thanks to a processing of 8,881 Landsat ETM+ images. Three different forest classes, namely the primary and secondary forests, and the woodlands, were obtained. The multi-date land-cover maps (i.e., for years 2000, 2005, and 2010) extracted from the original FACET data were resampled at 1-km in order to fit the spatial resolution of the population (Section 4.2.2) and the travel-time (Section 4.2.3) maps. The maps resampling did not have a significant impact on the amount of changes in the forest classes. Indeed, forest cover losses between 2000-2005 from the original FACET data set were 2,012 km^2 in woodlands, 3,666 km^2 in primary forest, and 11,681 km^2 in secondary forest (Table 1 in Potapov *et al.* (2012)). The corresponding estimates from the 1-km maps were respectively, 1,971 km^2 (increase of 2% by comparison to the 60-m data set), 3,701 km^2 (decrease of 0.95%), and 11,852 km^2 (decrease of 1.46%).

The total forest cover area for the year 2005 was estimated at 1,577,286 km^2, which represents approximately 67% of the country area. The forest cover losses between the 2005-2010 interval were 3,285 km^2 in woodlands, 7,007 km^2 in primary forest, and 9,468 km^2 in secondary forest. These losses can be considered such as minor in comparison to respective land-cover classes for 2005. Indeed, they correspond to 0.9% of woodlands area in 2005, 0.67% of primary forest area, and 5.53% of secondary forest area. The annual gross forest cover loss between years 2005 and 2010 was estimated at 0.25% of the 2005 forest cover area (Potapov *et al.*, 2012).

The land-cover map for year 2010 was used for the model validation.

4.2.2 Population maps

The population maps for the t_1 and the t_2 periods were obtained from the model developed in the Chapter 2. The corresponding population densities are averaged values of 30 simulations, which used the confidence intervals provided for each of the population density classes. For each simulated population map, a pixel-based random selection of density values inside the provided confidence intervals was achieved . The densities classes were assumed to have the same range of values for the two time periods. Population totals for years 2000 and 2005 were obtained from the FAO database[2], which used urban population growth rates ranging from 4.72% to 3.29%. Corresponding rates for rural population were comprised between 2.04% and 0.35% along the simulation period.

4.2.3 Travel time map

The model developed in the Chapter 3 was used to compute the travel time map for the whole DRC (Figure 3.9). Up to 165 major towns and big villages were included into the multimodal accessibility model such as destinations. These were selected based on a threshold of their total population in 2005 (i.e., $\geq$ 50,000 inhabitants). They were considered such as important populated places that drive the production of both agricultural products and charcoal, as well as the collection of fuelwood in surrounding rural areas.

The road network was obtained from the World Ressources Institute (WRI)[3] and the Université catholique de Louvain (UCL) - Geomatics database (Vancutsem *et al.*, 2007). The total length of the road network used was about 326,157 km. Approximately 17% of these, which included the roads found in urban areas, were considered such as in good condition, 28% in medium condition, and the remaining 55% were not suitable for car transportation. These latter included foot trails and logging roads. Logging roads and man-made trails such as those found in old industrial plantations were included since they were expected to play a major role in opening the forest blocks for the small holder agriculture (Makana and Thomas, 2006).

Additionally, some network sections were extended (up to 30 km outside the country limits) to integrate the DRC transportation network and major cities

[2] `http://faostat3.fao.org/home/index.html` - Accessed 19 February 2013
[3] `www.wri.org/publication/interactive-forest-atlas-democratic-republic-of-congo` - Accessed 28 February 2013

into the central Africa regional corridors (Gun Wang *et al.*, 2009). The western corridor considers the connection between Brazzaville (Republic of Congo) and Kinshasa, the DRC capital. This connection is important since it allows the DRC having a direct access to the deepwater port of Pointe Noire. The southwestern corridor is the railroad connection between the Katanga province (passing by Dilolo) and the Lobito port in Angola. The southern is dictated by the access to Durban, from Lubumbashi (in the Katanga province), passing by Lusaka (Zambia), Harare (Zimbabwe), and Johannesburg (Republic of South Africa). Finally, the eastern corridor is the connection between the city of Butembo and the Mombasa port, crossing Uganda and Kenya. The road condition of these connections was considered as suitable for car transportation.

The model also integrated the river network and ferries to build a fully modeled multimodal network. The corresponding data sets were obtained from the RGC[4] and the UCL-Geomatics database.

4.2.4 Protected areas

DRC protected areas (PAs) were obtained from the OFAC[5]. The PAs included into the developed model were the DRC's national parks that are listed such as the World Heritage Sites. These are: Salonga, Virunga, Kahuzi-Biega, and Garamba. Other PAs that are known to be under law enforcement were also included. This was for example the case for the Luki Biosphere Reserve in the Bas-Congo province, and the Lomako Faunal Reserve in northern DRC.

4.3 Methodology

The forest cover change simulations were run on annual time step and spanned over thirty years, starting from 2005. Three types of forest cover change were modeled and corresponded to forest cover transitions found between the t_0 and t_1 reference maps (Figure 4.1 - *Step II*), which were the conversions of both primary and secondary forests and woodlands to non forest areas. The model did not consider the forest regeneration nor the reforestation since these were not available in the reference maps used.

The developed model consisted of five main steps (Figure 4.1) that were the derivation of multi-date land-cover maps from the original FACET LCLCC map, the amount of change quantification, the potential for change computation, the spatial allocation of change, and the model validation. These steps are detailed in the following paragraphs.

[4] `http://www.rgc.cd/site/` - Accessed 28 February 2013.
[5] `http://www.observatoire-comifac.net/` - Accessed 28 February 2013.

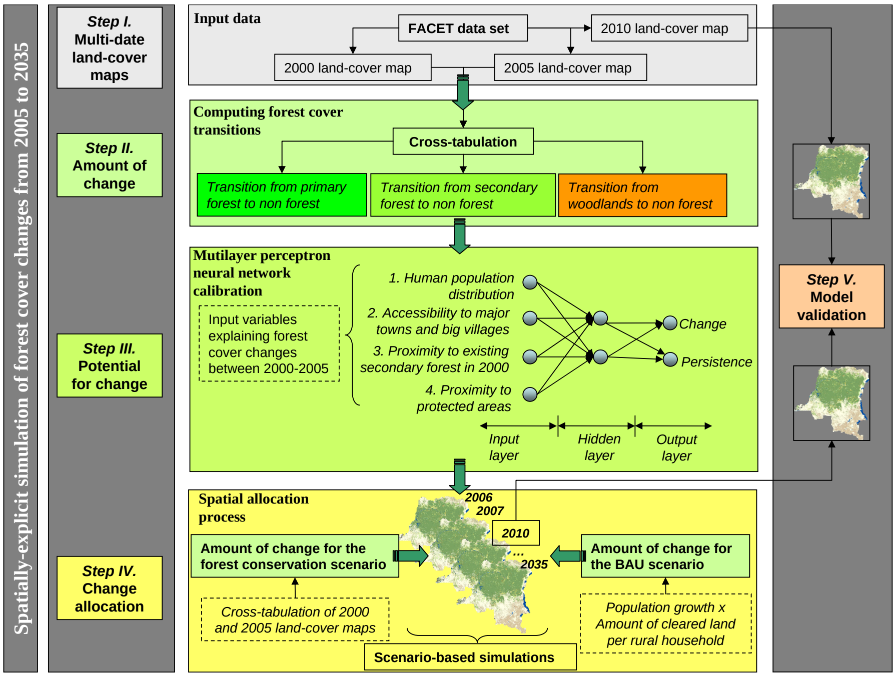


Figure 4.1: Five steps of the forest cover change model.

4.3.1 Amount of change

The amount of change for the forest conservation scenario was obtained from land-cover maps of years 2000 ($= t_0$) and 2005 ($= t_1$), which were derived from the FACET data set (Potapov *et al.*, 2012). Cross-tabulating these two reference maps (Figure 4.1 - *Step II*) allowed obtaining the amount of change for each type of forest cover loss, which were mapped such as conversions from primary/secondary forests and woodlands to non forest areas (Potapov *et al.*, 2012). For the forest conservation scenario, the amount of change corresponding to each type of forest cover loss during the period 2000-2005 was further used to simulate the forest cover changes up to 2035 (Figure 4.1 - *Step IV*).

For the BAU scenario, in turn, the total amount of change was computed using population projections obtained from the FAO database[6] and the quantity of cleared land needed per head of rural household, which was assumed to be about 0.25 ha (Tollens, 2010). Given the lack of precise information, it was also assumed that 30% of the rural population was involved in the clearing of new lands and a rural household consisted of six persons seeking for more fertile forested areas each five years. This assumption is close to the reality since using these model parameters with FAO estimates of rural population in 2005, i.e., 38,985,000, yield a total of 29,238.75 km^2 of forest cover loss between 2005-

[6]`http://faostat3.fao.org/home/index.html` - Accessed 19 February 2013

2010, which is of same magnitude than estimates from Potapov *et al.* (2012) for the same period. Linking the quantity of change with the population was also consistent with findings from Ernst *et al.* (2013) and Defourny *et al.* (2011) who indicated that the amount of forest cover change in DRC was correlated with the population density.

In addition, the population projections from FAO indicated that the DRC population will double in 2035 according to estimates of annual population growth rate. The population doubling within a period of thirty years seemed to be reasonable since it was already observed between 1984 — the DRC census with a total population of 30,731,000 inhabitants (Ngondo *et al.*, 1992) — and 2005, with a total population of about 60,000,000 of inhabitants (De Saint Moulin, 2006).

4.3.2 Potential for change

A potential for change was computed for each forest cover transition, using a multi-layer perceptron neural network (MLPNN) (Eastman *et al.*, 2005) (Figure 4.1 - *Step III*). It was chosen in this study based on its performance by comparison to other prominent procedures of determining the potential for change, such as the logistic regressions or the weights-of-evidence (Eastman *et al.*, 2005). The MLPNN was implemented using the Land Change Modeler module embedded in the IDRISI Taiga software.

4.3.2.1 Input variables

Four variables were selected for feeding the MLPNN. These were the spatial distribution of the human population (HPD) (Chapter 2), the output from the travel-time model (TTM) (Chapter 3), the proximity to both existing secondary forests (PSF) and protected areas (PAs). The HPD was used since it was assumed to be linked to the amount of cleared lands, mainly through shifting cultivation in rural areas for supplying urban markets and fuelwood collection - charcoal production for ensuring domestic energy needs (Tollens, 2010; Ernst *et al.*, 2013). It was also observed that the majority (61.98%) of forest cover changes during the period 2000-2005 were found in the lowest population density class (0-6 inhabitants/km^2) according to the model developed in the Chapter 2. This density class corresponds to land-cover types such as the dense moist forest, old and young secondary forests, savanna-woodlands, and tree savanna (Vancutsem *et al.*, 2008). 19.98% of these changes occurred within the class 74-218 inhabitants/km^2 that corresponds to rural complex areas, while 14.68% were found in the density-class 17-33 inhabitants/km^2, which refers to savanna areas. Only few changes (2.54%) were found within the class 135-514 inhabitants/km^2 and 0.77% within the class 48-139 inhabitants/km^2. The highest population density class corresponded to the lowest amount of change i.e., 0.04%.

Results from the TTM (Chapter 3) served as proxy-information for the accessibility to major towns and big villages which can allow approximating the patterns of forest cover changes. Using the modeled travel-time corresponds to a more realistic functioning of land-use change in DRC according to actual accessibility to populated places unlike classical proximity to infrastructures. Indeed, the single proximity to transportation networks can correspond to small villages that are at the same time far away from any major town where agricultural products, fuelwood, and charcoal are needed. As a consequence, the road/river proximity cannot fully explain land-cover changes. This was confirmed by the fact that about 70% of forest cover loss between 2000 and 2005 were found at locations lying within a daily journey from and to the nearest major town or big village, i.e., about 7 hours. Such observation clearly shows the added-value to account for a more realistic travel-time model which can link land-uses to territorial dynamics including specifically the accessibility to markets and the resulting small-scale economy.

The PSF was choosen according to the fact that secondary forests are mainly found along the road network and around villages that generally correspond to agricultural areas (Vancutsem *et al.*, 2008), which can have a direct impact on the forest cover changes. 81.53% of changes between 2000-2005 were found at less than 1km of existing secondary forest, 12% between 1-3km, 2.09% between 3-5km, and 4.38% at more than 5km. The PAs were used since they often act such as an effective tool to reduce the forest cover loss (Clark *et al.*, 2008; Ernst *et al.*, 2013). This was consistent with findings from Potapov *et al.* (2012) who indicated that forest cover loss in DRC national parks was lower than the national average.

These four variables (i.e., HPD, TTM, PSF, and PAs) were therefore assumed in the present research to correctly approximate the drivers of forest cover change in DRC, such as already identified by Defourny *et al.* (2011) who used the estimates of forest cover changes provided by Ernst *et al.* (2013). The possible correlation between these variables was tested using the Pearson's coefficient.

4.3.2.2 Multilayer perceptron neural network

The MLP is the most widely used ANN that consists — in the developed model — of three layers which are the input (i.e., the predictors of change), the output (i.e., the expected transition), and two *hidden* layers (Hinton and Salakhutdinov, 2006) (Figure 4.1 - *Step III*). Each of these layers contain a series of interconnected nodes. Except for the input layer, the nodes are processing elements (i.e., neurons) that use activation functions in order to define the output of each node given an input or a set of inputs (Khoi and Murayama, 2011).

The number of nodes in the input layer (N_i) corresponds to the four used explanatory variables. The output layer has two nodes (N_o) corresponding to the two possible states: change (i.e., state = 1) or persistence (i.e., state = 2). The number of nodes in the *hidden* layer N_h is computed according to the following Expression 4.1 (Eastman, 2009):

$$N_h = INTEGER(\sqrt{N_i \times N_o}) \tag{4.1}$$

Changed pixels between t_0 and t_1 belonging to the forest transition to be modeled are randomly assigned to both the training (calibration) and the testing sets. The training process aims at obtaining the proper weights for both the connection between the input and the *hidden* layers, and between this latter and the output layer, in order to correctly classify the pixels such as changed or unchanged.

The first step of the MLPNN calibration involves the training set and consists into the identification of the n variables that drive the targeted transition. Each of these variables (i.e., $X = x_1, x_2, \ldots, x_n$) is associated with a neuron in the input layer which standardizes the x_i values between -1 to 1 (Eastman, 2009).

In the second step, the standardized variables (i.e, x'_i) are distributed to each of the neurons in the *hidden* layer and are multiplied by a weight ($\omega_{i,j}$). The role of the *hidden* layer nodes, in combination with the connection weights, is to create a set of hyperplanar discriminant functions that can separate the variable space (i.e., the driving forces of forest cover change in this case) into regions associated with the dependent variables (i.e., forest cover transitions). In the *hidden* layer, the signal received by the neuron j for a pixel k (i.e., $\sigma_j(k)$) is calculated as follows (Expression 4.2) (Eastman, 2009):

$$\sigma_j(k) = \sum_i \omega_{i,j} x'_i(k) \tag{4.2}$$

The resulting values in the neurons of the *hidden* layer are again distributed to each neuron of the output layer during the third step of the MLPNN. The function (i.e., the activation function) that indicates the transition probability (Pr) for the neuron l in the output layer is most often a sigmoid that represents the non-linearity of each node (Eastman *et al.*, 2005):

$$Pr(k, l = 1) = \sum_j \omega_{j,l} \frac{1}{1 + e^{\sigma_j(k)}} \tag{4.3}$$

where possible values for l are 1 or 2 corresponding respectively to the change or the persistence.

At the completion of the three first steps (i.e., the *forward-propagation* process), the likelihood of the weights assigned to each predictor of change in the output layer is assessed using the testing set of pixels. The assessment method, which is called the *back-propagation* of the errors, is based on a relatively simple concept. If the network gives the wrong answer, then the weights are corrected to reduce future errors. The error resulting from the comparison between pixels with high potential for change and changed pixels from the testing set is then spread backward to correct the weights set. The weights' adjustment is achieved through the application of the *delta rule* (Eastman *et al.*, 2005) (Expression 4.4):

$$\Delta\omega_{i,j}(m+1) = \eta(\delta_i\sigma_j) + \alpha\Delta\omega_{i,j}(m) \tag{4.4}$$

where $\Delta\omega_{i,j}$ is the adjusted weight, m is the number of iterations (i.e., 10,000 for the developed model), η is the learning rate parameter, δ_i is an index of the error change rate, and α is the momentum parameter. For the presented model, this latter was set to its default value (i.e., 0.5), while the learning rate yielded good results for all transitions at the value 0.001.

The corrected weights are introduced into a new forward process during which the training set is presented to the network iteratively, until the error of the network is minimized or reaches an acceptable magnitude. This is evaluated based on the root mean square error (RMSE) (Eastman, 2009):

$$RMSE = \sqrt{\sum_{p=1}^{P}\sum_{k=1}^{N}(T_{pk} - O_{pk})^2} \tag{4.5}$$

with P being the number of classes corresponding to the possible states, which are the change and the persistence in the developed model. N is the number of pixels in the training set, T is the state resulting from the model and O is the actual state. If the network is trained successfully, the RMSE typically declines along the training process until a stabilization at its lower level. The network training is however unsatisfactory if the error stops dropping, or starts to rise.

Finally, an accuracy rate (AR) is computed (Expression 4.6) for each transition using the neural network model exhibiting the lowest RMSE. The accuracy rate measures the ability of the neural network to correctly classify the pixels belonging to the testing set. It is therefore expressed such as the percentage of the total number of testing pixels that are correctly classified such as changed or unchanged by the model.

$$AR = \left(\frac{\sum_{k=1}^{M}(T_kO_k)_{state=1} + \sum_{k=1}^{M}(T_kO_k)_{(state=0)}}{M}\right) \times 100 \tag{4.6}$$

where M is the total number of pixels in the testing set.

4.3.3 Change allocation

For each forest cover transition in both scenarios, locations of possible future forest cover losses were predicted based on the corresponding potential for change maps (Figure 4.1 - *Step IV*). The allocation was achieved by selecting the pixels exhibiting highest potential for change values, up to reaching the required amount of change. This method is justified in the developed model since there was no conflicting allocation. Indeed, the transitions from primary forest to non forest only occurred in primary forest, and the same for the two others transition types.

4.3.4 Model validation

The land-cover map of year 2010 ($= t_2$) was used to validate the simulated land-cover map resulting from the BAU scenario (Figure 4.1 - *Step V*). The overall agreement between both maps was computed to highlight the quantity (dis-)agreement between the simulation and the original map.

In addition, location (dis-)agreements were evaluated using the histogram method (Eastman, 2009). Accordingly, for each transition, changed and unchanged pixels between t_1 and t_2 reference maps were summarized based on their potential for change values. A transition modeled with a good performance will exhibit a bi-modal-like pattern, i.e., most of unchanged pixels between t_1 and t_2 will correspond to low potential for change values, while changed pixels will have high values.

4.4 Results

4.4.1 Total amount of change

The transitions from forest classes to non forest areas between 2000-2005 are presented in the Table 4.1. The total area of change corresponds to 1,971 km^2 (11.25%) in woodlands, 3,701 km^2 (21.12%) in primary forest, and 11,852 km^2 (67.63%) in secondary forest. These values will therefore correspond to the amount change per transition for simulating the patterns of change in the future according to the forest conservation scenario. The resulting forest cover area in 2035 was estimated at 1,487,118 km^2, which represents 63% of the country area. Forest cover loss areas between 2005-2035 were estimated at 11,593 km^2 (12% of the total forest cover loss area) in woodlands, 21,957 km^2 (24%) in primary forest, and 56,618 km^2 (62.79%) in secondary forest, which adds up to 90,168 km^2. These forest cover loss estimates correspond respectively to 3.17%, 2.14%, and 33.24% of the classes area in 2005.

Table 4.1: Markovian transition probability matrix between years 2000 and 2005. Corresponding areas in thousand km^2 are between brackets.

	2005					
2000	**NF**	**ND**	**WT**	**WO**	**PF**	**SF**
Non Forest (NF)	$1_{(687.456)}$	NA	NA	NA	NA	NA
No Data in DRC (ND)	0	$1_{(45.44)}$	NA	NA	NA	NA
Water (WT)	0	0	$1_{(10.31)}$	NA	NA	NA
Woodlands (WO)	$0.0054_{(1.97)}$	0	0	$0.9946_{(365.72)}$	NA	NA
Primary Forest (PF)	$0.0035_{(3.70)}$	0	0	0	$0.9965_{(1,040.62)}$	NA
Secondary Forest (SF)	$0.0648_{(11.85)}$	0	0	0	0	$0.9352_{(170.95)}$

The forest cover area in 2035 corresponding to the BAU scenario was estimated at 1,364,066 km^2, representing 58% of the country area. The resulting total forest cover loss area was about 213,220 km^2, corresponding to 29,843 km^2 (14% of the total forest cover loss area) in woodlands, 59,731 km^2 (28%) in primary forest, and 123,646 km^2 (58%) in secondary forest.

4.4.2 MLPNN estimations

The correlation between the spatial variables used to construct the change potential maps was tested using the Pearson's coefficient (Table 4.2). The highest value (-0.49) was found between the travel time to major towns and the distance to secondary forests. It can therefore be concluded that no significant correlation was found between the variables used.

Table 4.2: Pearson's coefficients between the spatial variables.

	HPD	TTM	PSF	PAs
HPD	1.00000	0.00402	-0.04354	-0.00258
TTM	0.00402	1.00000	-0.49422	0.29951
PSF	-0.04354	-0.49422	1.00000	-0.04285
PAs	-0.00258	0.29951	-0.04285	1.00000

The best variables combination for each transition is presented in the Table 4.3, along with the corresponding accuracy rate. The transition from primary forest to non forest exhibited the highest accuracy rate (i.e., 79%) followed in a decreasing order by the conversions of woodlands (i.e., 58%) and secondary forest (i.e., 54%) into non forest areas. The interpretation of the variables' weights is, however, hampered since they cannot be considered such as the variables contribution in explaining the variability in the transition modeled. In turn, these weights can be interpreted such as the relative importance of each variable by comparison to the others for each transition.

Table 4.3: Weights assigned to each spatial variable for the different transitions.

	Weights of N_i to N_h nodes		
	PF ⇒ NF	**SF ⇒ NF**	**WO ⇒ NF**
TTM	-1.4660	-0.1998	-1.0067
PSF	-8.3015	–	–
PAs	-0.0892	0.1275	0.1724
Accuracy rate	**79%**	**54%**	**58%**

The simulated map for the year 2010 under the BAU scenario was compared to the t_2 reference map for highlighting location agreement and disagreements. Mostly due to the fact that changes between years 2005 and 2010 for the three transitions were minor by comparison to the respective land-cover classes areas in 2005 (Section 4.2.2), the overall agreement between the two maps (Table 4.4) was high and reached 98%. A full map of location agreements and disagreements is provided in Figure 4.2. Selected location disagreements (Figures 4.2a to d) are enlarged on Figures 4.3 and 4.4 and are discussed in detail in Section 4.5

Table 4.4: Overall agreement between classes from the reference map and the simulated map under the BAU for the year 2010. Areas are provided in km^2.

	BAU simulation for 2010						
FACET 2010	NF	WT	WO	PF	SF	Total	Agreement
Non-Forest (NF)	705,987	0	3,256	6,755	8,549	724,547	97.44%
Water (WT)	0	45,442	0	0	0	45,442	100%
Woodlands (WO)	4,043	0	358,370	0	0	362,413	98.88%
Primary Forest (PF)	8,002	0	0	1,025,641	0	1,033,643	99.23%
Secondary Forest (SF)	16,223	0	0	0	145,440	161,663	89.96%
Total	734,255	45,442	361,626	1,032,396	153,989	2,327,708	–
Agreement	96.15%	100%	99.10%	99.35%	94.45%	–	97.99%

Location agreement and disagreements were also evaluated using the histogram method (Section 4.3.4). The histogram referring to the 'primary forest ⇒ non forest' transition (Figure 4.5b) illustrates that most of changes in primary forest between 2005-2010 were found at locations corresponding to high potential for change values. Indeed, it was observed that about 84% of pixels transitioning from primary forest to non forest had their potential for change value higher than the median value (i.e., 0.462). The histogram corresponding to pixels remaining in the primary forest class between 2005-2010 (Figure 4.5a) shows that they corresponded to lower potential for change values. 63% of these pixels had a potential for change value lower than or equal to the median potential for change (0.406). It can therefore be concluded that the potential for change corresponding to the transition 'primary forest ⇒ non forest' performed well in locating changed and unchanged pixels between 2005-2010.

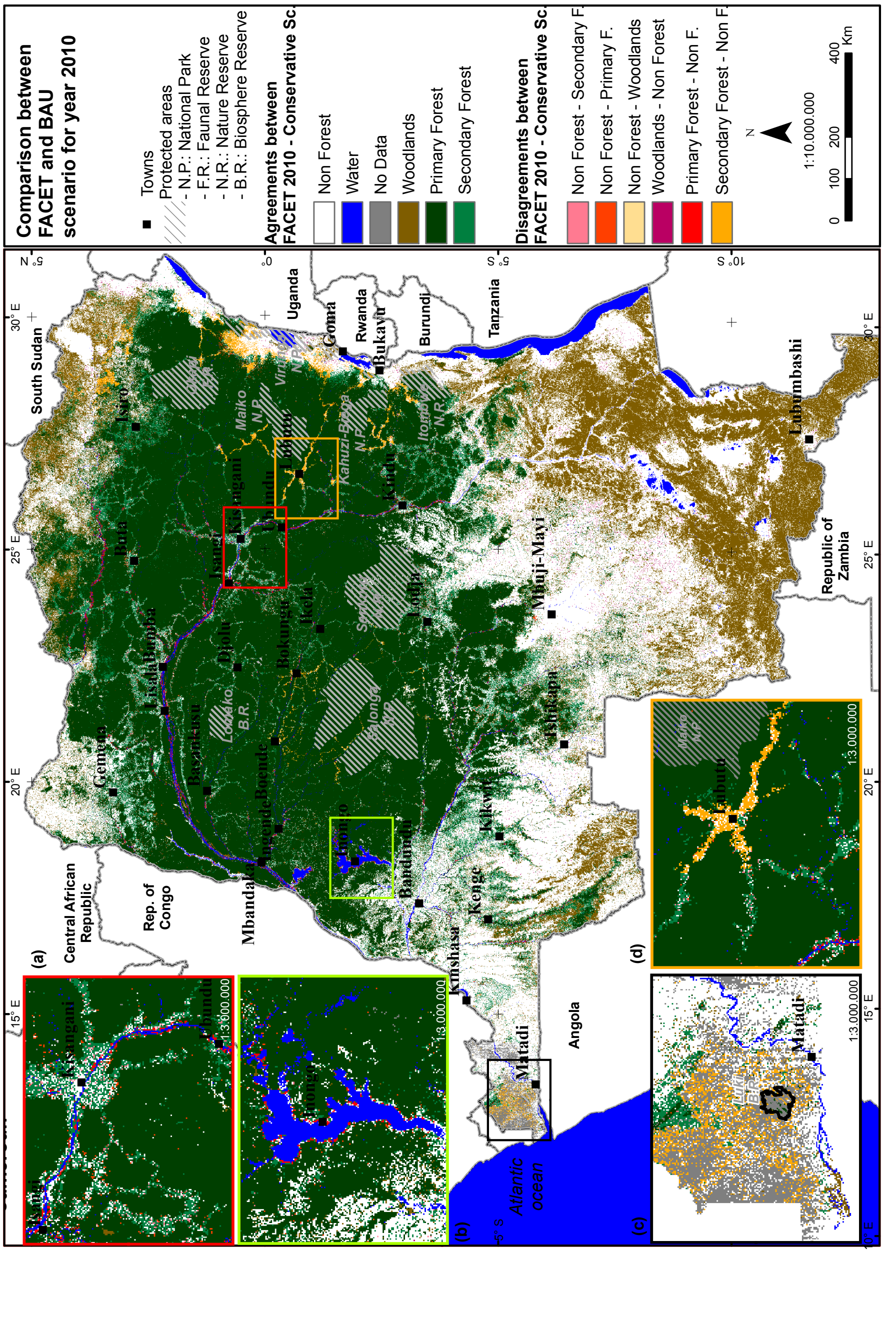


Figure 4.2: Comparison between the reference map and the BAU scenario simulated map for year 2010.

For the transition 'secondary forest ⇒ non forest', about 79% of changed pixels between 2005-2010 had their potential for change higher than the median value (i.e., 0.472) (Figure 4.5d). However, only 18% of pixels remaining in the secondary forest class between 2005 and 2010 had their potential for change lower than or equal to the median value (0.454) (Figures 4.5c) . In turn, about 58% of pixels transitioning from woodlands to non forest for the period 2005-2010 had a potential for change value higher than the median value (i.e., 0.484) (Figure 4.5f). 45% of unchanged pixels between the two years for the woodland class had a potential for change lower than or equal to the median value (0.466) (Figure 4.5e). Therefore, potential for change models for both secondary forest and woodlands had a lower performance to locate corresponding changed and unchanged areas between 2005-2010. These low performances are discussed in the Section 4.5.

4.4.3 Predicted land-cover maps from the two scenarios

Outputs from both scenarios of years 2010, 2015, 2020, 2025, 2030, and 2035 are presented in Figures 4.6, 4.7, and 4.8 for a selected area located between Kisangani and Basankusu downwards the Congo river, within the congolese *Cuvette centrale*. It can be observed that edaphic forests found along the Congo river, those in the North of Basankusu, and those found along main rivers such as the Tshuapa river (Boende - Bokungu - Ikela axis) are not affected by changes. These areas have indeed low potentials for change. These low values are close with the reality since these frequently-inundated areas are found on poorly drained soils (Vancutsem *et al.*, 2008), which do not allow any agricultural activity.

The few changes observed for areas along the Boende - Bokungu - Ikela axis are also consistent with findings from Potapov *et al.* (2012) which showed that forest cover changes in these zones were minor for the period 2000-2005 and 2005-2010. This can be explained by the fact these three towns are located in the middle of the Cuvette Centrale and are therefore poorly connected to important cities such as Kisangani or Mbandaka. In addition, the DRC Territoires corresponding to these three towns have also very low population densities (De Saint Moulin, 2006). As a consequence, forest cover loss predictions according to both scenarios were quite similar along the Boende - Bokungu - Ikela axis, although minor differences were found for the first ten years of simulation. These differences are likely to indicate that the amount of change obtained from the population growth (i.e., BAU scenario) was overestimated in comparison to the amount of change obtained from the period 2000-2005 for this specific region. From the simulation of years 2015 till 2035, the amount of change from both scenarios become close, leading to similar patterns of forest cover change.

Major location disagreements were observed for the Lisala-Bumba region, along the axis Kisangani - Buta and in neighborhoods of these two cities, and around Basankusu and Djolu. Predicted forest changes mostly corresponded to these major towns footprints, which is a direct consequence of using a travel-

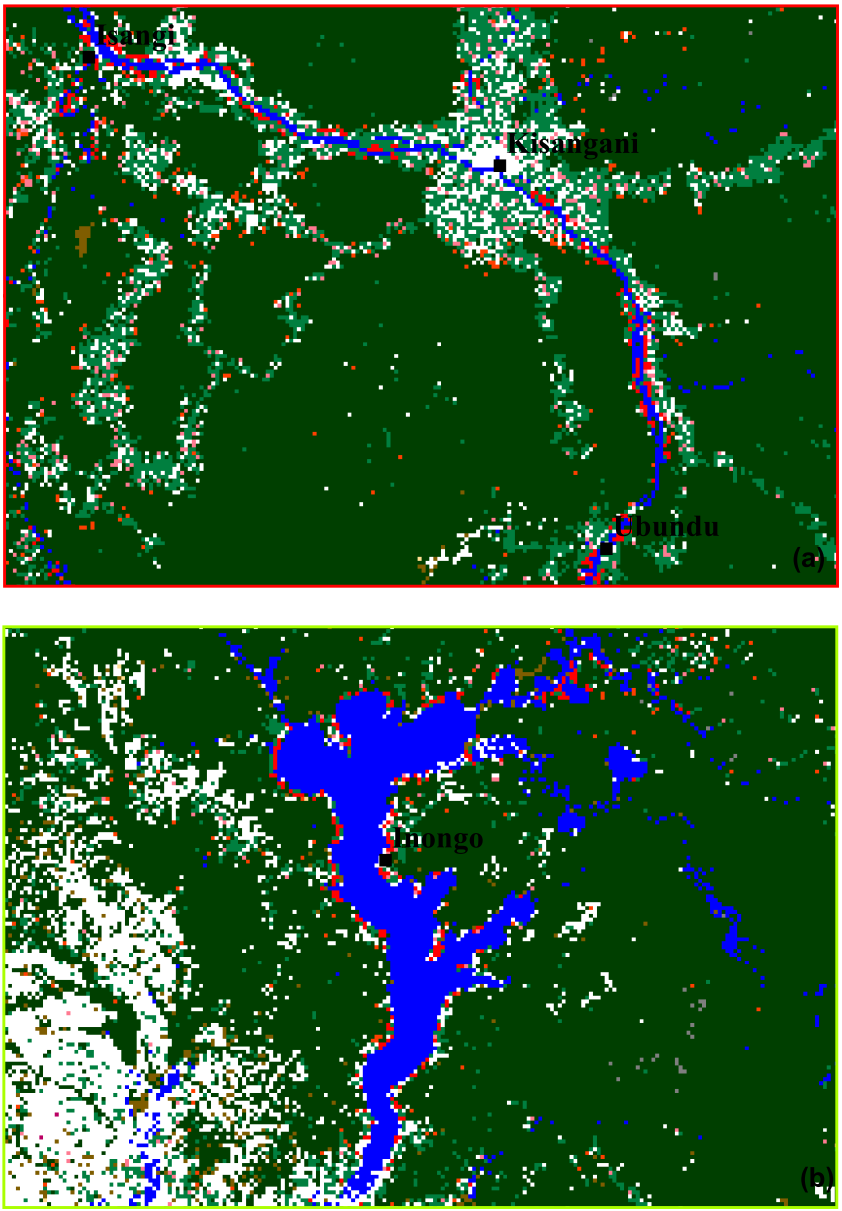


Figure 4.3: Disagreements between reference map and BAU scenario in 2010 for (a) Kisangani and (b) Inongo areas. Colors correspond to the legend provided in Figure 4.2

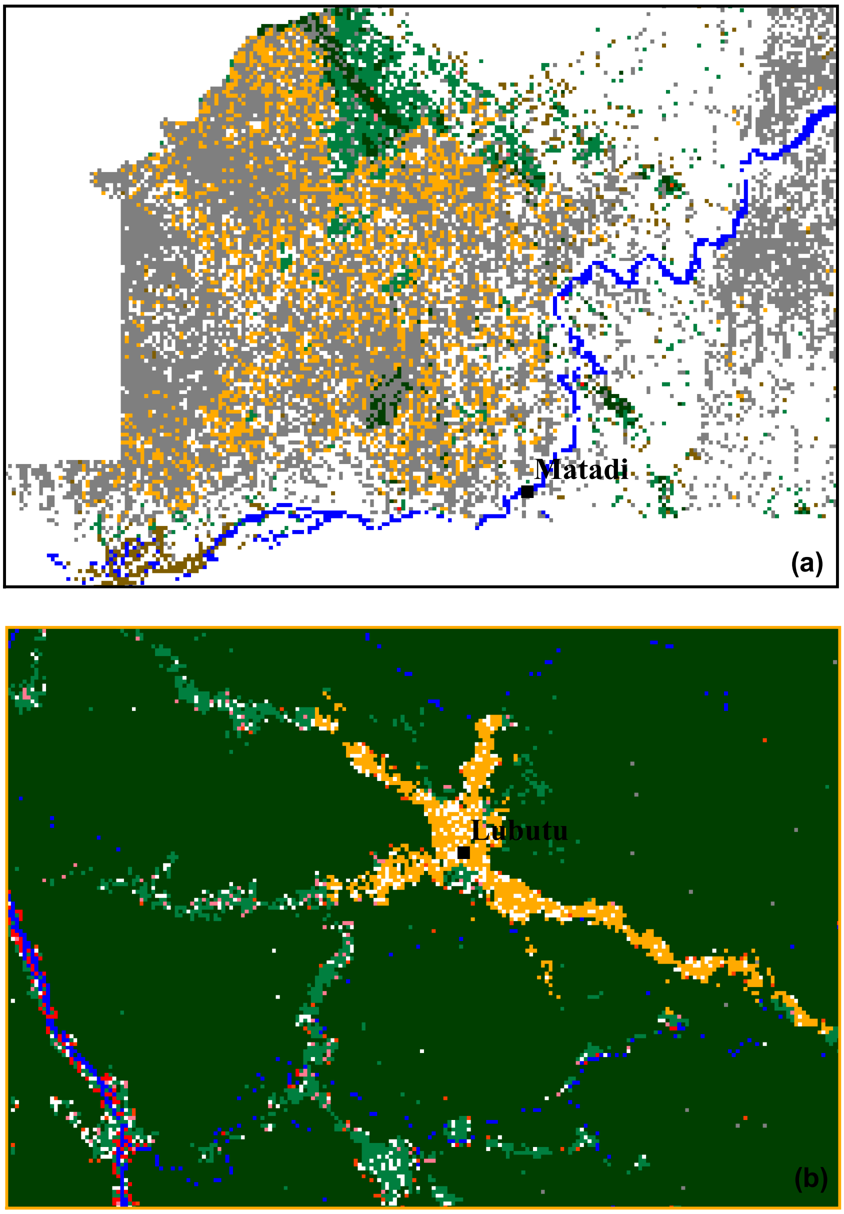


Figure 4.4: Disagreements between reference map and BAU scenario in 2010 for (a) Matadi and (b) Lubutu areas. Colors correspond to the legend provided in Figure 4.2

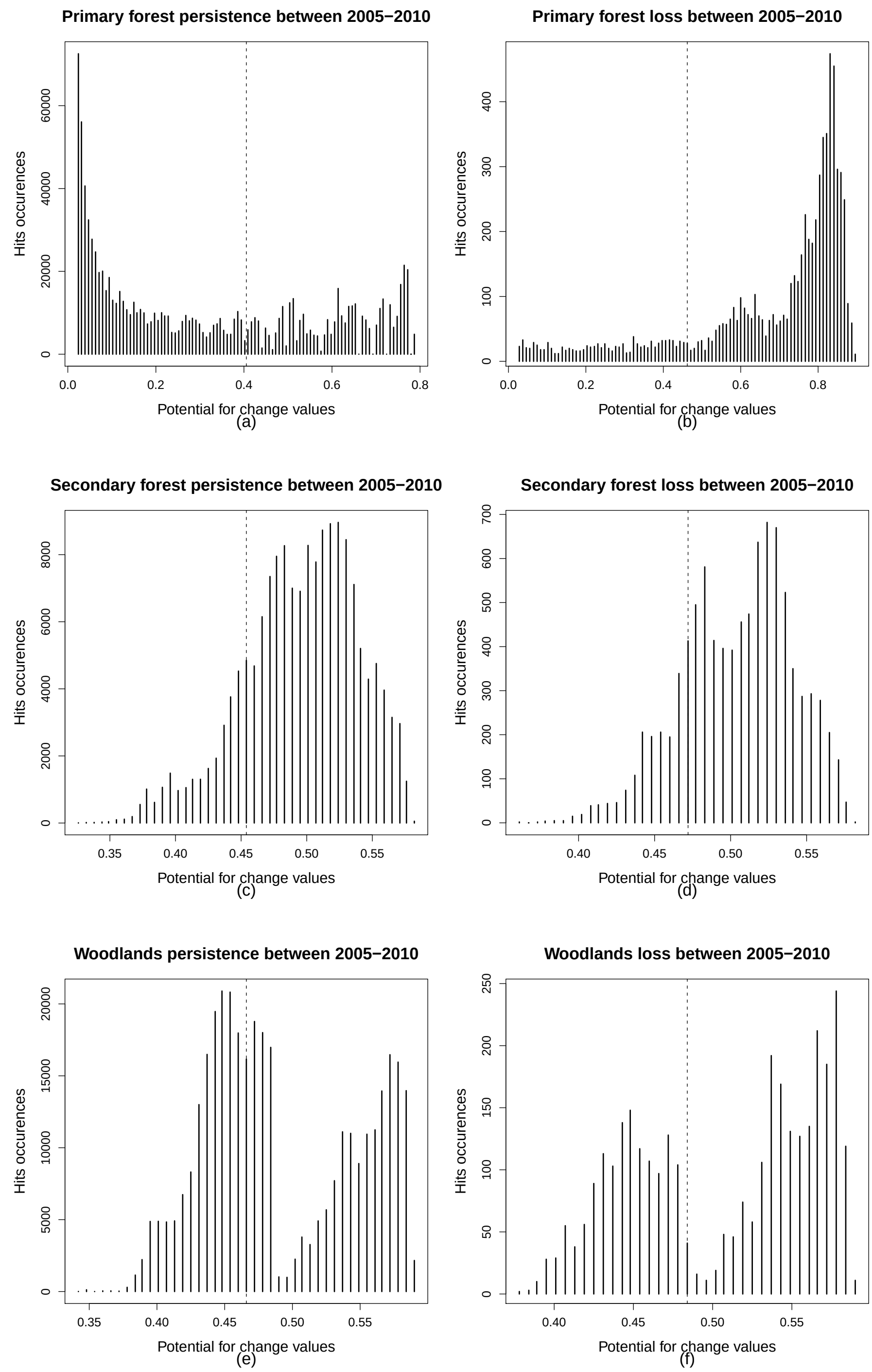


Figure 4.5: Model performance for each type of forest cover loss. The dashed line represents the median value of the potential for change per transition.

time variable to model forest cover changes. A cost-distance-based variable is unlikely to reproduce such patterns of change, since these would mostly correspond to buffered areas along the road/river network.

Two distinct patterns of forest cover change dynamics were found within the *Cuvette centrale* from both scenarios (Figures 4.6, 4.7, and 4.8). The first corresponds to important forest fragmentation in the northern part due to the proximity to major towns and big villages, i.e., Basankusu, Djolu, and Isangi. The second pattern was found in the middle of the Cuvette (i.e., in the region of the Salonga National park) corresponding to forests that will likely remain undisturbed or very few disturbed in the next years.

4.5 Discussion

The developed model was based on a set of spatial variables that were assumed to approximate the drivers of forest cover change at the DRC national scale. These variables were the human population distribution, the accessibility to major towns and big villages, the proximity to existing secondary forests in 2005, and the proximity to protected areas.

The selection of changes' predictors was a critical step for the overall modeling process. Some previous studies (Nelson and Geoghegan, 2002; Verburg *et al.*, 2004a) concluded that the predictors of land-cover changes are most often not independent. This was also argued in favor of using artificial intelligence methods since they can account for correlated variables (Sangermano *et al.*, 2010). The present research used, however, few and uncorrelated variables. The clear advantage of using such variables is to reduce the risk of model over-fitting.

According to selected variables, it was observed that 61.98% of forest cover changes which occurred between 2000 and 2005 were found in the lowest population density class (0-6 inhabitants/km^2), which corresponds to land-cover classes such as dense moist forest, old and young secondary forests, and savanna-woodlands (Vancutsem *et al.*, 2008). About 20% of these changes were found in the density class corresponding to rural complex areas (i.e., 74-218 inhabitants/ km^2). Less than 1% of changes was found in urban areas.

In addition, 70% of changes between 2000-2005 corresponded to locations that were at 7 hours of the nearest major town or big village, which correspond to a daily journey to and from a populated place. This finding demonstrates the distinct advantage of using the multimodal travel-time model instead of previous distance-based proximity to roads/rivers used in most, if not all, land-cover change model. A more realistic travel-time modeling such as achieved in this thesis allows approximating adequately the land-cover change dynamics in DRC. In fact, forest cover changes in DRC are mostly linked to subsistence activities relying on small-scale economy which includes the accessibility to

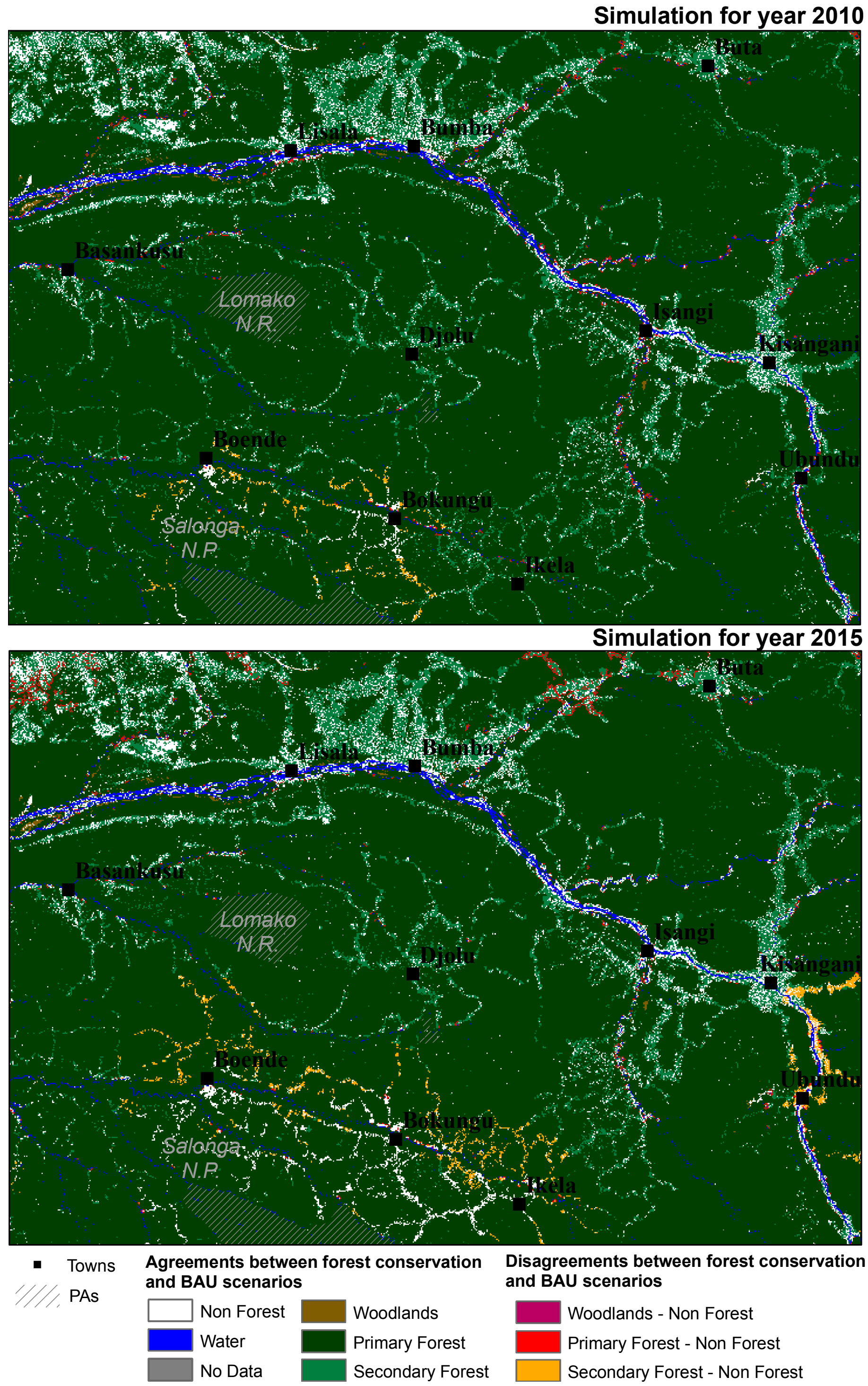


Figure 4.6: Simulated forest cover losses from both scenarios for years 2010 and 2015.

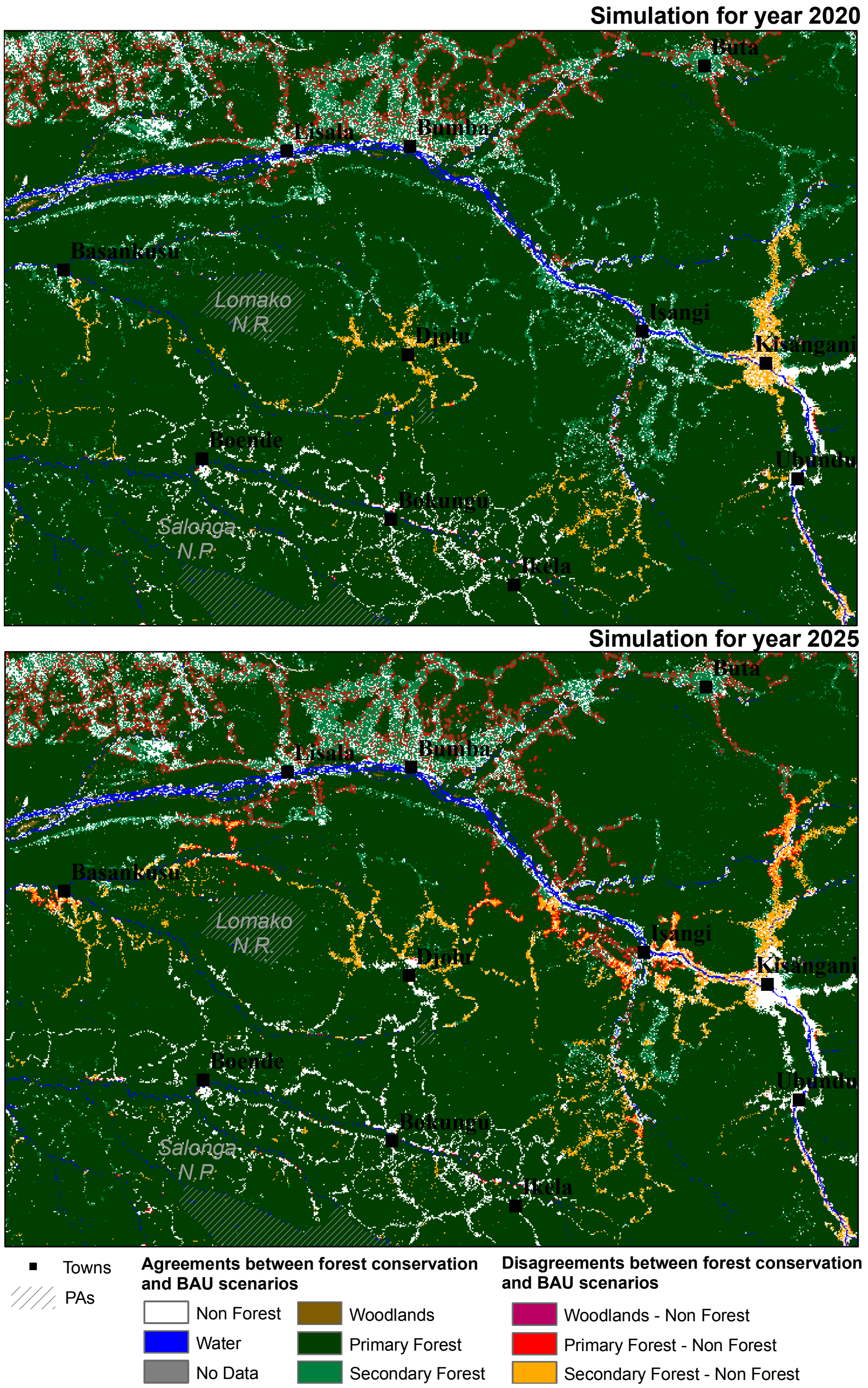


Figure 4.7: Simulated forest cover losses from both scenarios for years 2020 and 2025.

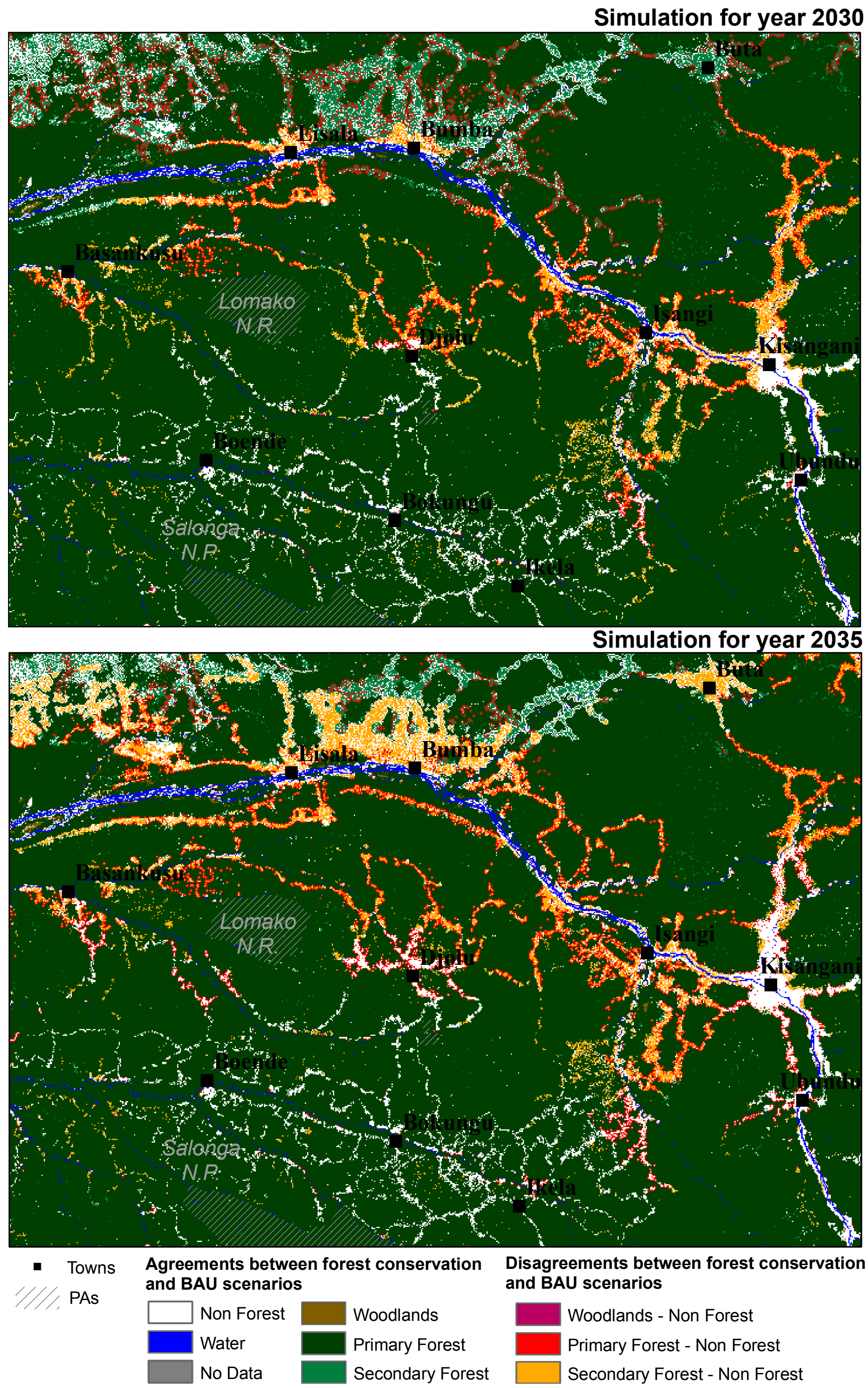


Figure 4.8: Simulated forest cover losses from both scenarios for years 2030 and 2035.

markets and major towns, where agricultural products, fuelwood and charcoal are needed. A single proximity to a road network does not provide a consistent information about the accessibility to markets since villages along DRC roads can be far away from any market or major town. The advantage of using a multimodal travel-time as a predictor of forest changes was also illustrated by the change patterns which mostly corresponded to major towns and big villages' footprints, while commonly used distance to roads/rivers cannot depict such patterns.

Also, the accessibility to major towns and big villages had an impact on the three transitions modeled (Table 4.3), while the human population distribution did not. This finding supports that the prediction of the amount of forest cover change should be modeled separately than the prediction of the changes' locations. Indeed, in countries such as the DRC, the internal demand for forest products is mainly driven by the household energy needs and the shifting cultivation, which are linked to the total amount of people or the population increase. The location of forest changes is, in turn, mainly linked to the proximity to major towns. Furthermore, since the demand for household energy is high in urban areas, estimating the spatial distribution of the urban and rural populations — such as achieved in the Chapter 2 — is therefore relevant.

The variable distance to secondary forest was used based on the analysis of forest cover changes in primary forests between 2000 and 2005. Since secondary forest mainly corresponds to rural complex zones (Chapter 1), forest blocks in the immediate proximity of secondary forest areas are more likely to be under greater human-induced pressure. This variable was used in conjunction with the accessibility to major tows and protected areas, which yielded a a good model performance for the transition 'primary forest ⇒ non forest' (i.e, Accuracy rate = 79%).

Besides, the medium performance of the model 'woodlands ⇒ non forest' transition (i.e., Accuracy rate = 58%) can be due to the fact that the changes in woodlands are more likely to be clustered around southern DRC towns (Figure 4.9a) or dispersed in northern savanna areas. The fact that woodlands are distributed either side of the DRC dense moist forest extent could indicate that a different type of predictors may be more suitable for these zones, or a geographically-stratified model could be used to monitor these specific changes. Indeed, change dynamics occurring in savannas such as grazing and fires are different from changes in dense moist forests (Zida *et al.*, 2007).

The model had a lower performance for the transition 'secondary forest ⇒ non forest' (i.e., Accuracy rate = 54%). This can be linked to the specific dynamics occurring within the secondary forest, that are mainly related to the shifting cultivation. Changes observed within the secondary forest class are indeed made of many isolated pixels representing a 'peper-and-salt' pattern within rural complex areas (Figures 4.9b, c, and d).

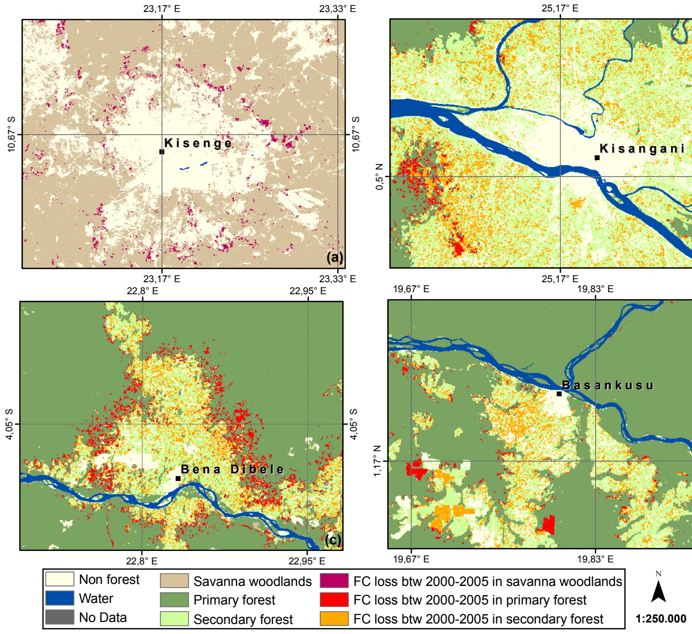


Figure 4.9: Patterns of forest cover loss for selected locations from the original 60-m FACET land-cover and land-cover change map.

The simulated land-cover map for the year 2010 under the BAU scenario exhibited a high overall agreement (i.e., 98% in Table 4.4) in comparison to the t_2 reference map, since changes between 2005-2010 in the different forest classes were minor (Section 4.2.1). However, this indicates that the BAU scenario is a good short-term predictor of future forest cover changes in the DRC. The spatial comparison between both maps allowed highlighting location disagreements. A total area of 3,256 km^2 (Table 4.4) which changed from woodlands to non forest between 2005-2010 were missed by the model. These missed change locations correspond to single pixels scattered in southern DRC. 6,755 km^2 that transitioned from primary to non forest during the same time-period were predicted such as remaining in this land-cover class. They also correspond to single isolated pixels that were mainly found at the edges of the DRC forest extent. Also, the developed model did not capture the transition of 8,549 km^2 of secondary forest to non forest. These were broadly co-located with wrong predictions of primary forest persistence. In addition, the model predicted 10,223 km^2 of secondary forest loss while these areas did not change according to the t_2 reference map. This could be due to the potential for change overestimation, since these false alarms of change were mainly found in the neighborhood of major towns such as Bukavu, Goma, and Isiro (Figure 4.2), and along the road network linking Bukavu to Lubutu (Figures 4.2d and 4.4b).

Other areas of major disagreements were found along rivers and lakes' banks. This was the case along the Congo river, from Ubundu downstream to Yangambi, passing by Kisangani (Figure 4.2a and 4.3a). The same was observed along the banks of the Maï-Ndombe lake (Figure 4.2b and 4.3b). While these locations could correspond to fishing camps (or similar human settlements) which could indeed induce forest cover changes (Zhang *et al.*, 2006), the (quasi-) systematic over-predictions of forest cover losses in these zones should be put in question. A reason of such over-predictions could be the important weight assigned to the travel-time model in such areas that are close to navigable rivers or lakes. It was also observed that 'No Data' areas (mainly due to cloud contamination) in the reference maps had an impact on the location of predicted forest cover losses. This is likely to be the reason of the changes' over-prediction in the western DRC (Bas-Congo Province) (Figure 4.2c and 4.4a). The presence of important cities such as Tshela, Lukula, and Matadi in this area coupled with a quite dense road network led to high values of potential for change. Accordingly, the model will then tend to convert the major part of forest pixels into non forest.

Results from both scenarios for the period 2005-2035 show that the population increase may induce forest cover changes that seriously depart from the forest conservation scenario trends (Table 4.5), particularly in secondary forest. This was corroborated by Zhang *et al.* (2006) who indicated that under current technology level, the demand for agricultural land in central Africa may more than double in the next 20 years (i.e., around 2025) solely by accounting for the population growth. The Table 4.5 illustrates the forest cover loss differences between both scenarios. For each simulation period, the area corresponding to the forest conservation scenario is deducted from the corresponding area for the BAU scenario. The difference is reported to the class area in 2005. The larger differences are found for the secondary forest class.

Table 4.5: Comparison of forest cover loss areas between both scenarios.

Simulation period	Primary forest	Secondary forest	Woodlands
2005 - 2010	0.44%	3.44%	0.58%
2010 - 2015	0.50%	4.80%	0.70%
2015 - 2020	0.58%	6.09%	0.79%
2020 - 2025	0.65%	7.25%	0.89%
2025 - 2030	0.69%	8.24%	0.97%
2030 - 2035	0.77%	9.39%	1.06%

Estimates of forest cover change obtained from Zhang *et al.* (2002); Potapov *et al.* (2012); Ernst *et al.* (2013) allows speculating on the model results. For the forest conservation scenario, the annual rate of change between 2005-2035 corresponds to 0.20%. This rate seems unrealistic, without land-use policies, in comparison to trends observed between 1990-2010. Indeed, Ernst *et al.* (2013) observed an increase of the annual net deforestation and degradation rates from respectively 0.11% and 0.06% to 0.22% and 0.12% between 1990-2000 and 2000-2005 for DRC. These estimates correspond to an increase of ratio 1:2 between the two periods. In addition, Potapov *et al.* (2012) found an increase in the annual forest cover loss from 0.22% for 2000-2005 to 0.25% for 2005-2010. That being said, the low rate of forest cover loss resulting from the forest conservation scenario tends to indicate that the DRC will likely remain a high forested and low deforested country if mitigation actions, i.e., REDD+ projects, are undertaken to reduce the forest cover loss.

For the BAU scenario, the annual rate of forest cover loss between 2005-2035 was estimated at 0.48%, almost two and a half times greater than the rate resulting from the forest conservation scenario. The corresponding estimation from Zhang *et al.* (2002) is about 1.2%, which is six times larger than the forest conversation scenario rate. Tollens (2010) found the rate obtained by Zhang *et al.* (2002) to be overestimated since the rural-urban out-migration was not taken into account, and the different model parameters (i.e., shifting cultivation of 15 years, rural population growth above 3%, and rural population density of about 25 inhabitants/km^2) were quite overvalued. The developed BAU scenario, in turn, used a rural population growth ranging from 0.35% to about 2% [7], while the maximum value of the population density in forest was 6 inhabitants/km^2 (Chapter 2). However, the developed model did not consider fallows, neither possible reforestation or forest regeneration since these types of forest changes were not distinguished in the reference maps used and were pooled in the secondary forest class.

To further analyze the BAU scenario outcome at the end of the simulation period, patterns of change between the t_1 reference map was compared to the simulated map in 2035 (Figure 4.10). Main changes were found in eastern DRC, in areas that are close to major towns such as Bumba, Buta, Kisangani, and Isiro in North DRC, Mbuji-Mayi, Kananga in the South. Kisangani will likely play an important role since this city connects the *Cuvette centrale* to western DRC through the Congo river, to northern savanna regions through Buta and Isiro, to eastern DRC through Bunia and Walikale-Bukavu, and to South DRC through Kindu. These connections were all prioritized in the planning of infrastructure building currently conducted by the DRC Government. Impact assessment of infrastructure building on the forest cover is therefore highly required.

Change dynamics in northeastern DRC correspond to forest-savanna transitions, particularly in the North-East of the Okapi wildlife reserve limits. These areas also correspond to the axis Isiro-Buta, which are highly populated towns connected to both Kisangani and northeastern DRC up to Uganda. Moreover, the 'triangle' Isiro-Buta-Kisangani is considered such as an important asset for developing the entire Province Orientale. Consequently, DRC forests in the immediate proximity of this region will likely be under increasing pressure in the next years. Forest cover losses in eastern DRC will mainly originate from population pressure in the Kivus and the connection between this part of the country to Kisangani. Due to the persistent political unrest in eastern DRC since twenty years, population migration and refugee camps will likely still have

[7] `http://faostat3.fao.org/home/index.html` - Accessed 19 February 2013

a negative impact on the forest resources. This is particularly critical for nature conservation since this region hosts some of the major national parks (Virunga, Kahuzi-Biega, Maïko) and nature reserves such as Tayna and Itombwe.

Change dynamics in more central DRC refer to areas that are along the Congo river, particularly the Gemena-Lisala-Bumba axis. This region was formerly one of the most important center of industrial agriculture (e.g., coffee, palm oil, rubber, cotton) in the country, which still mainly driven by agricultural activities. The proximity of the Congo river is an incentive for such activity, since it is potentially the cheapest means of transportation for supplying bigger markets such as Kinshasa, Mbandaka, and Kisangani. Another type of change dynamics was found along the Congo river, downwards from Kisangani to Mbandaka, passing by Basankusu. Main activities are agriculture, fishing, fuelwood and charcoal production. In addition, the transportation network found in the *Cuvette centrale* which connects cities such as Djolu, Isangi, Ikela, Bokungu, Boende, Ingende, and Mbandaka will likely increase the forest fragmentation.

Change dynamics in southern DRC — although less marked in the Figure 4.10 — are either linked to agriculture (Lodja region) or industrial and artisanal extractive activities in areas that are rich in metal deposits (Mbuji-Mayi, Tshikapa). Finally, dry forests in the Bandundu region (Kenge, Kikwit, Bandundu) will likely face drastic changes mainly due to the fact that this region is both highly populated and a major provider of agricultural products for Kinshasa, the DRC capital.

The developed forest conservation scenario indicated that the DRC will likely remain a high forested and low deforested country if mitigation actions are undertaken to reduce the forest cover changes. This is certainly a key statement for DRC negotiators involved in the diverse international initiatives aiming at reducing the deforestation and the forest degradation in developing countries. The DRC has therefore to be supported for developing appropriate forest management policies, while financial incentives such as those framed in the REDD+ initiative are an important component to ensure that the country is compensated for its contribution to global efforts in mitigating climate change. Meanwhile, the BAU scenario indicated that the population doubling in 2035 will also double the annual rate of change obtained from the forest conservation scenario for the period 2005-2035.

In addition, the infrastructure development currently planned by the DRC Governement will likely increase the risk of forest change. Impact assessment of road development and rehabilitation must therefore be conducted since the existing transportation network was likely contributing to forest fragmentation, according to the developed model. The DRC therefore needs to develop a strategic land-use planning at the national scale in order to conciliate the economic development with a sustainable use of the forest resources. Such land-use planing should involve decision makers from diverse sectors (i.e., forest manage-

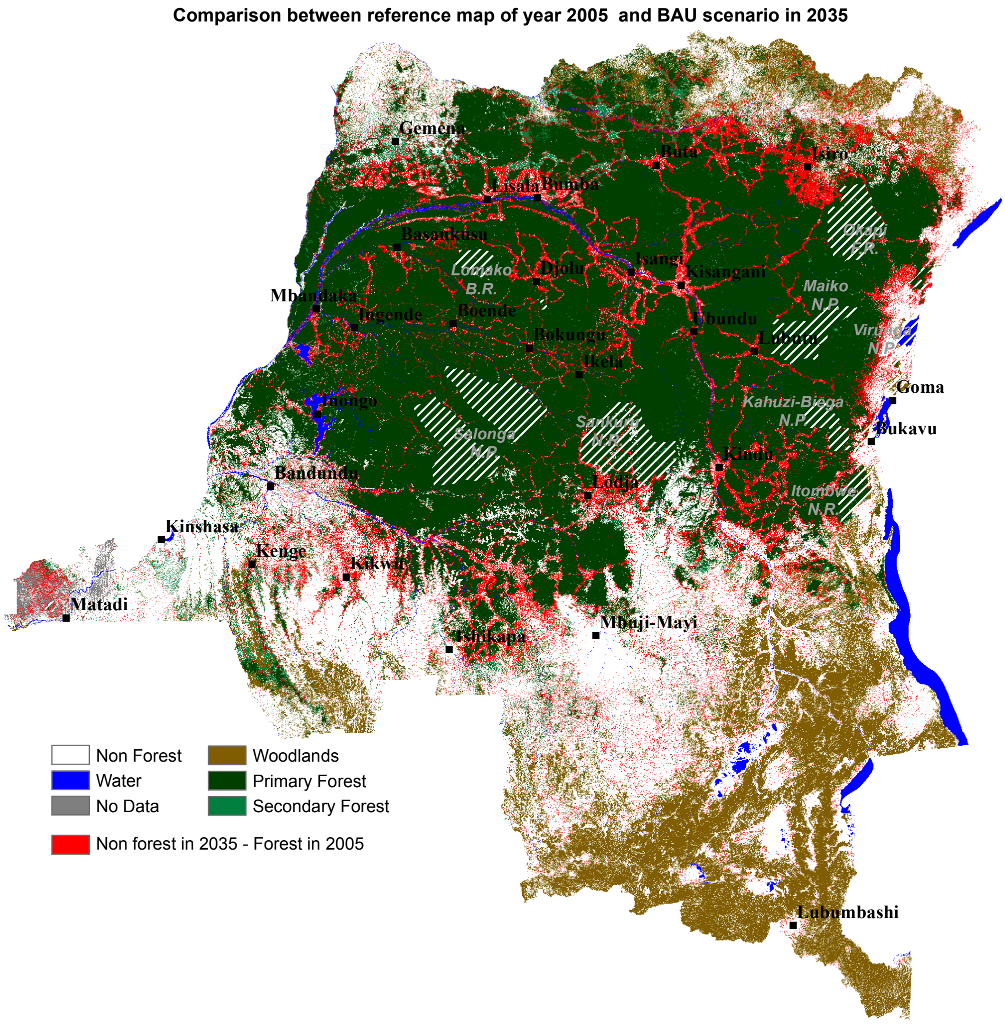


Figure 4.10: Comparison between reference map of year 2005 and BAU scenario in 2035.

ment, agriculture, mining, nature conservation) since changes dynamics resulting from both scenarios depicted diverse patterns of change across the country. These were likely originate from various activities, such as mining, charcoal production, and subsistence agriculture. For this latter in particular, there is a crucial need to promote agricultural practices ensuring better crop yields in a sustainable manner, while also contributing to maintain an acceptable level of soil fertility over time. Such practices would limit the progression of agricultural frontiers into intact forest blocks such as observed from predicted forest cover changes between 2005 and 2035 (Figure 4.10). The national land-use planning has also to be envisioned in the regional context of Congo basin countries, which also committed to mitigate forest cover changes (Tadoum *et al.*, 2012), as well as in the international context since (future) forest cover changes in the DRC may be driven by the increasing international demand for both meat and biofuel (Megevand *et al.*, 2013).

4.6 Conclusion

The drivers-based model developed in this chapter used the information from two reference land-cover maps to simulate the future forest cover changes in the DRC for the next thirty years, starting from 2005. The potential for change was constructed using a multi-layer perceptron neural network, which was trained with the back-propagation of errors method.

At the difference of previous modeling efforts using MLPNN for modeling the forest cover changes in the tropics (Mas *et al.*, 2004; Sangermano *et al.*, 2010; Khoi and Murayama, 2011), the model implemented in this Chapter used few and uncorrelated variables, which prevent any model over-fitting. In addition, the variables used were directly meaningful such as illustrated by the results from the multimodal travel-time model developed in the Chapter 3. Indeed, it was found that 70% of changes for the period 2000-2005 were located at a daily journey from and to the nearest major town or big village.

The BAU scenario was found to be a good short-term predictor of forest cover change, since forest cover losses observed between 2000-2005 were minor compared to forest classes' areas. As a consequence, land-cover classes resulting from the simulated map of year 2010 exhibited a high overall agreement (i.e., 98%) in comparison to the 2010 land-cover map. However, except for the transitions in primary forest, the model did not predict accurately the locations of change in secondary forest and woodlands. These are, indeed, hard to predict from a nationwide modeling since they occur at small scale (shifting cultivation in rural complex areas) or in specific areas (southern and northern DRC savannas). For these latter, using a stratified model which can take into account the specific dynamics occurring in savanna could be helpful.

A possible approach to account for local-scale forest cover change dynamics could be coupling agent-based models with cellular automata simulations. This kind of models could benefit from participatory mapping exercises (Colom, 2006; Nackoney *et al.*, 2013) that can allow linking the dynamics of agricultural land-use to the location of villages, and include statistics such the distances from human settlements to the agricultural limits, the agricultural land used per person, the fallow lengths, the ratio of the agricultural activities in the household income, etc.

The developed model allowed exploring the results from two distinct scenarios. Yet, the lack of precise information that could allow building more elaborated assumptions has limited the exploration of other types of simulations. Results from more oriented socio-economic models, such as global (partial) equilibrium models (Schneider *et al.*, 2011), can be accounted for to provide estimates of forest changes in the future. Though, outputs of these kind models are generally available at coarse spatial resolution (i.e., 50-km or more). Spatially-explicit modeling such as developed in this Chapter can thus be coupled with these economic-oriented model for the change allocation.

The forest conservation scenario indicated that the DRC will likely remain a high forested and low deforested country by implementing mitigation actions (i.e., improving forest management, promoting sustainable agriculture, conducting impact assessment of roads rehabilitation and construction) to reduce the forest cover changes. The BAU scenario was found to be a good short-term predictor of forest cover change, while it showed also that the DRC population will at least double in 2035, resulting in doubling the rate of change estimated from the forest conservation scenario. The spatial comparison of both scenarios highlighted some contrasted forest cover change dynamics over the country, i.e., in the Lisala-Bumba region, in eastern DRC, and within the *Cuvette centrale*. A national land-use planing is therefore highly required so as to allow conciliating the economic development with a sustainable use of the forest resources.

Conclusion and perspectives

The overall objective of this thesis was to investigate the possible futures of the Democratic Republic of the Congo forests by developing a spatially-explicit and dynamic model of forest cover loss at the national scale. This main goal was subdivided into three research questions that structure a detailed presentation of the thesis main findings in the following section.

Findings

Can the results from independent remote sensing methods (i.e., pixel- and object-based classifications) for monitoring forest cover changes be combined in order to provide consolidated estimates?

Chapter 1 developed an analysis framework which demonstrated how the pixel-based information could be accounted for at the object-level, in order to achieve a comparison between the two approaches. Such comparison aimed at building the compatibility and possibly the consistency of forest cover change estimates obtained from independent image processing methods using the same type of satellite imagery.

After discussing the methodological differences between pixel- and object-based approaches using the most recent products delivered for the DRC, a detailed cross-comparison analysis of their respective land-cover classes was conducted for years 2000 and 2005. From this comparison, the main classes of the pixel- and object-based products, respectively the primary forest and the tree cover ones, exhibited an agreement of 90%. Total forest cover area estimates computed from the two products turned out to be quite close, i.e., 1,102,930 km^2 and 1,012,000 km^2 respectively for pixel- and object-based products for the year 2000. For the year 2005, these estimates were respectively 1,088,270 km^2 and 998,530 km^2. Objects delineating pixel-based secondary forests mostly corresponded to the no tree cover class according to the object-based approach, which considered these areas such as rural complex zones. The disagreement secondary forest - rural complex area resulted mainly from the choice of the minimum object size in heterogeneous land-cover classes that are a mix of fallows, vegetation regrowth, food crops areas, agriculture areas in forest, and small holder agriculture. Indeed, at the scale of the minimum object size of 5 ha, as implemented in the object-based approach, the percentage of tree cover within the objects in these specific zones could be in fact lower than 10% of the object size, leading to consider these areas as belonging to the object-based no tree cover class. This disagreement had a direct impact on the change estimates from the two approaches since most of the forest cover loss (i.e., 67%) during the study period occurred within the secondary forest. The present thesis completed a systematic evaluation of the discrepancies between secondary forest and rural complex classes from visual interpretation of Landsat images for selected sites. Though, further assessment based on spectral information from high resolution images is needed to strengthen the distinction between these two land-cover types. Also, object-based mosaic classes corresponded to a mix of the three forest classes (i.e., 40% of primary forest, 40% of secondary forest, and 10% of woodlands) in the pixel-based product. Although mosaic classes are of particular interest to depict change dynamics in the forest-savanna and forest-non forest frontiers, the distinction between high and low mosaics based on different forest cover proportions within the objects could not be achieved successfully using the information from the pixel-based product. Grouping these two classes into a single mosaic class may provide more consistent information for the further comparison with the pixel-based product.

The analysis also demonstrated how shifting cultivation, which is one of the main causes of forest cover loss in the DRC, could be handled efficiently by working at the object level. A pixel-based approach will likely consider these temporary trees removal as gross deforestation, leading to a risk of overestimating the deforestation rate. Object-based products also facilitate the forest cover change accuracy assessment since analysts can directly interpret homogeneously delineated objects, while such validation is less straightforward for pixel-based products. Moreover, working at the object-level allows mapping the rate of forest cover changes to spatialize this indicator of the pressure on the forest resources. Such rate of change mapping cannot be achieved using pixel-based products.

Finally, the extensive visual interpretation of discrepancies between pixel- and object-based products allowed deriving consolidated annual rates of deforestation (0.29% ± 0.05%), forest degradation (0.16% ± 0.03%), forest regeneration (0.08% ± 0.02%), and reforestation (0.04% ± 0.01%). Such rates of change derived from independent remote sensing methods are critical information for decision makers. They indeed contribute to the credibility (by demonstrating that estimates of forest cover change from different approaches are quite close) and the transparency (by conducting a detailed comparison analysis of two main satellite images processing methods) of forest monitoring from satellite Earth observation for DRC. The developed consolidation approach is also relevant for the REDD+ initiative in DRC, since country-based estimates of forest change have to meet some level of international standards in terms of precision, transparency, repeatability, and uncertainty evaluation, prior any financial allocation rewarding the avoided deforestation/degradation.

Can spatial variables quantifying the main driving forces of the forest cover loss be appropriately retrieved in a poor data context of the DRC?

Obtaining spatial variables quantifying the forest cover change drivers was an essential requirement to fulfil the overall objective of this thesis. Retrieving these variables was, in turn, hindered by the scarcity of the available information in the poor data environment of DRC. The variables of interest for quantification and mapping were the human population distribution and the accessibility to major towns and big villages. These two variables were expected to impact both the quantity and the patterns of forest cover change at the national scale.

The human population distribution for DRC was modeled (Chapter 2) based on available coarse scale population counts (i.e., at the third country administrative level) in conjunction with a nationwide scale land-cover map at 1-km spatial resolution to obtain population densities per land-cover type. Population densities in urban areas were found to be the highest (3,526 - 5,260 inhabitants/km^2), followed by intensive and extensive agricultural areas (135 - 514 and 74 - 218 inhabitants/km^2), mountain forests (48 - 139 inhabitants/km^2), and savanna areas (17 - 33 inhabitants/km^2). The lowest population density was found for classes such as dense moist forest, forest-savanna mosaic, and savanna-woodlands (0 - 6 inhabitants/km^2). The confidence intervals for each population density class illustrated the distinct advantage of the implemented model since available global and continental population data sets (i.e., GRUMP, LandScan, and Afripop) only deliver a single population density value, without attaching any reliability information about the estimate provided. The comparison between population estimates from most common population distribution models and those obtained from the developed model showed that this latter poorly performed (SSE = 3,073,349) to fit locally (i.e., at the third country administrative level) the population totals. This was not surprising since these global/continental population data sets are constrained by the total population at the source zone scale without assessing the reliability of these local estimates. In the specific case of the DRC, available population counts at the third country administrative level are merely projections from the last population census done in 1984. Meanwhile, the range of plausible population densities per land-cover type obtained in this thesis can be used to compute a set of population distribution maps from which it is afterwards possible to select the one minimizing the differences at the local scale.

Chapter 3 aimed at computing the travel time to selected destinations that was used such as proxy information for the accessibility to major towns. The developed multimodal accessibility model optimized the use of the existing transportation network (roads, railroads, and rivers) in conjunction with alternative transportation systems (*off-road* network), which were all weighted by the topography to obtain the minimum travel time to the nearest destination from any location within the study area.

The multimodal accessibility model demonstrated its realism when analyzing the output patterns in comparison to cost-distance-based models. This realism was exemplified using three specific real-world cases. The first case showed that the single proximity to a network component *per se* does not necessarily imply a reduced travel-time to the destination. Indeed, the travel-time to the destination for a location close to the network is a function of both the travel speed on the network and the friction surface between the actual location and the network component. The second case highlighted the use of *off-network* travels in conjunction with a multimodal transportation network, which can imply diverse transfers between available travel modes, in order to optimize the travel-time to the nearest destination. *Off-network* travels are traditional and are used in the absence of adequate transportation networks. The travel-time model also implemented the possible restrictions of transfer between specific network segments and modeled the transfer locations that are, for example, existing bridges or ferries. The last case referred to optimizing *off-network* travels for reaching the fastest network component in order to minimize the travel time to the destination, while distance-based models will only account for the nearest network segment without considering real-world restrictions of transfer between network components.

Can the retrieved spatial variables approximating the main forest cover change drivers for the DRC be jointly analyzed in order to simulate both the quantity and the patterns of past and future forest cover losses at the national scale?

Chapter 4 used a neural network-based model fed by the results obtained in the three previous chapters for simulating the possible future forest cover changes in the DRC for the next thirty years, starting from 2005. The model had the distinct advantage of using only few (i.e., four) directly meaningful and uncorrelated variables to reduce the risk of model over-fitting. Two scenarios were implemented and the resulting outputs were compared to each other in terms of both the quantity and the patterns of simulated forest cover changes. The forest conservation scenario assumed that the amount of forest cover changes in the future will be of same magnitude such as observed during the period 2000-2005. Such scenario could result from land-use planning policies, possibly incentivized by REDD+ projects. The business-as-usual scenario showed that the DRC population will (at least) double in 2035 according to annual population growth rates. It was also assumed that a rate of 0.25 ha of cleared land is needed per head of rural household (Tollens, 2010) along the simulation period.

Results from the forest cover change model indicated that the human population distribution was not a predictor for the patterns of future forest changes. These patterns corresponded, however, to major town footprints thanks to a more realistic characterization of the accessibility to populated place obtained from the multimodal travel-time model. 70% of forest cover changes between 2000 and 2005 were found at locations lying within a daily journey from and to the nearest major town or big village, i.e., about 7 hours. Such finding cannot be obtained using classical (weighted-) distances to roads/that are used in most, if not all, land-cover change models. Using a more realistic travel-time model ensures to correctly approximate the land-cover change dynamics in DRC that are mainly linked to subsistence activities relying on small-scale economy and the accessibility to markets. A single proximity to roads and rivers cannot provide such information since villages along the transportation network in DRC can be far away from any major town where markets for agricultural products are located and where charcoal and fuelwood are used for the domestic energy needs.

The predicted amount of forest cover loss in 2035 according to the forest conservation scenario was about 5.72% of the 2005 forest area, resulting in annual change rate of 0.20% between the two years. These estimations were found to be unrealistic – without policy regulations - in comparison to the trends observed from 1990 to 2010. Indeed, Ernst *et al.* (2013) observed an increase of annual net deforestation and degradation rates from respectively 0.11% and 0.06% to 0.22% and 0.12% between 1990-2000 and 2000-2005. These estimates corresponded to an increase of ratio 1:2 between the two periods. In addition, Potapov *et al.* (2012) found an increase in the annual forest cover loss from 0.22% for 2000-2005 to 0.25% for 2005-2010. The population doubling between 2005 and 2035 according to the business-as-usual scenario resulted, in turn, into a forest cover loss of 13.52% of the 2005 forest area, corresponding to both an increase of 123,052 km^2 of forest cover loss area by comparison to the forest conservation scenario and a change rate of 0.48% between the two years. Furthermore, the patterns of forest cover change according to both scenarios were similar and mostly corresponded to footprints of major towns and big villages and to areas close to the transportation network by moving away from big towns. This also indicated that road construction and rehabilitation already planned by the DRC Government will likely increase the risk of forest cover loss, which suggests conducting quantitative and spatially-explicit impact assessment of infrastructure building using modeling approaches such as achieved in the present thesis.

The forest conservation scenario indicated that the DRC will remain a high forested and low deforested country if mitigation actions are undertaken to reduce the forest cover changes. The business-as-usual scenario was found to be a good short-term predictor of future changes, although the doubling of the population in 2035 will increase progressively the forest cover change rate by comparison to the forest conservation scenario outcomes. These two contrasted scenarios can be used as a basis for the ongoing REDD+ process in DRC, particularly for the construction of both the national reference level of emissions and the crediting baseline which would result from adjusting the reference level to the country national circumstances (Karsenty and Ongolo, 2012).

As a final word, both the human population distribution and the travel-time to major cities estimated respectively in Chapters 2 and 3 were proved to be appropriate spatial variables for simulating the amount and the patterns of future forest cover changes in DRC.

Perspectives

The present thesis aimed at contributing to the necessary quantitative assessment of the actual DRC forest resources and the possible future forest cover changes in regards to challenges of agriculture development, biodiversity, forest management, and land-use planning. These challenges have led the international community to put in place several initiatives for reducing the impacts of the human activities on the forest resources, especially in developing countries. Forest cover monitoring in these countries is however hindered by the lack of the necessary resources and the scarcity of available data. Although the approach developed in this thesis allowed successfully retrieving some of the critical spatially-explicit information needed to model the future forest cover changes in the DRC, the findings detailed in the previous section also highlighted some new challenges for research and further policy implications for a more efficient forest cover monitoring.

Next steps for the research

The quantitative assessment of the DRC forest cover over years 2000-2005 achieved by comparing two independent remote sensing methods highlighted the necessity of clearly distinguishing (young) secondary forests from rural complex areas. Such distinction has indeed a direct impact on the forest cover change estimates using multi-date land-cover maps. Although Verhegghen *et al.* (2012) aggregated these two land-cover types into the same class, their distinction was already discussed in Vancutsem *et al.* (2008), even if the corresponding mapping output at 1-km of spatial resolution (Vancutsem *et al.*, 2007) was too coarse to depict the small-scale forest cover changes lying within rural complex zones. Therefore, mapping these two classes at a spatial resolution comprised between 300 and 30 meters would allow both refining the estimates of forest cover and detecting some of the subtle forest cover changes commonly found in DRC. Indeed, this range of spatial resolution corresponds to a scale where natural and human-induced changes can be detected, differentiated, characterized and efficiently monitored over time for the whole country. Although a clear distinction between rural complex areas and secondary forests was not achieved for the year 2010 (Potapov *et al.*, 2012), it should be accounted for to deliver future forest cover change maps for DRC, possibly starting from 2015.

Another important challenge is the necessary distinction that has to be made between the deforestation and the temporarily tree removal resulting from shifting cultivation practices (Angelsen, 2001). Results from forest degradation monitoring using remote sensing-based canopy contiguity assessment and vertical stand structure coupled with field-based information (i.e., assessment of the growing woody biomass) (Baldauf *et al.*, 2009) have to further be compared to estimates provided by Ernst *et al.* (2013) and Duveiller *et al.* (2008), which are based on the tree cover proportion within elementary units of analysis.

Although spatial variables quantifying the major drivers of forest cover change in the poor data environment of DRC were successfully retrieved at 1-km spatial resolution, their implementation at the scale corresponding to small forest cover changes commonly found in DRC has to be considered. For the population distribution model in particular, applying the developed algorithm to land-cover maps of varying spatial resolutions – preferably lower than 1 km – and comparing the resulting population densities per land-cover type to those obtained in the present thesis would allow assessing the generalization of the constrained-regression method to model the population distribution. Also, coupling predicted land-cover maps such as those obtained in this thesis with population distribution maps derived from projected population totals could allow simulating the extension of existing settlements as well as the infrastructure building.

Meanwhile, since population densities per land-cover type obtained in the current research can only be validated using census-based information, further research is needed to (1) reliably characterize the original 1984 census data (e.g., number and size of enumeration units, census strategy, census error) and (2) estimate the error associated to the census-derived population projections obtained from De Saint Moulin (2006) and used in this thesis to approximate the spatial distribution of the human population.

The framework designed to simulate future forest cover changes included a mechanistic selection of pixels with highest potentials of change. As a consequence, each yearly-based forest cover change simulation between 2005 and 2035 only delivered a single output instead of a set of plausible land-cover maps. This mechanistic assumption could be relaxed by implementing a more stochastic-oriented selection of pixels lying within the range of high potential values. For each year, the resulting simulated maps can therefore be used to compute the final land-cover map, which could be constructed from pixels with the highest majority score estimated from the total number of occurrences during the diverse realizations of the stochastic selection.

Only one forest cover transition (i.e., from forest classes to non forest) was modeled in this research due to the few available multi-date land-cover maps for the DRC. Developing a new set of such land-cover maps at the national scale is therefore required for improving the simulations of future forest cover changes. These land-cover maps should be tailored to achieve (at least) the objectives of representing major biomes and depicting the land-cover classes that can facilitate the detection of the main forest cover change processes in DRC. Currently, the few existing land-cover maps do not achieve these objectives. The classes of particular interest could therefore be the following: primary forest, old secondary forest, young secondary forest, recently cleared land, forest-savanna mosaic, savanna, and the rural complex. Indeed, such land-cover classes would allow depicting the deforestation (i.e., conversion from primary and secondary forests to recently cleared lands), the forest regeneration (i.e., conversion from young secondary forest to old secondary forest or conversion from cleared land to young secondary forest), fallows (i.e., vegetation regrowth within both young secondary forests and recently cleared lands), shifting cultivation within both rural complex and recently cleared lands, etc. In addition, most of the proposed classes have the advantage to have quite distinct spectral signatures from the analysis of satellite imagery (Vancutsem *et al.*, 2008), which could facilitate the image classification process from both pixel- and object-based approaches.

Overall, the spatially-explicit simulation of future forest cover change implemented in this thesis was mainly data-driven oriented, which timely benefited from the few available multi-date land-cover maps for the model calibration and validation. However, implementing a process-driven model could also be an interesting approach for further research. The emphasis will therefore be on simulating the human activities such as shifting cultivation, fuelwood collection, charcoal production, mining, and illegal logging that have a direct impact on the forest resources. Since these processes most often interact complexly, such modeling is expected to be high data demanding and therefore quite challenging to be implemented in the context of DRC. A more oriented process-driven model should also consider the demand for forest resources originating from international markets. Therefore, coupling both data- and process-driven models could lead to spatially allocate the proportion of each process and retrieve the corresponding ratio in the total amount of change computed from multi-date land-cover maps.

Policy implications of findings

The present thesis highlighted the lack of reliable and consistent information at the DRC national scale for an efficient forest cover monitoring. This observation is of paramount importance since the DRC voluntarily committed to participate into the REDD+ initiative, which is a performance-based process that necessitates the most precise estimates of forest cover changes prior to any payments for the avoided deforestation/forest degradation. As a consequence, a more formal institutional support to national entities in charge of land-use planning, forest management, country statistics, cartography, and nature conservation is highly required from national decision makers. Such support would lead to provide these congolese institutions with the necessary resources to allow them collecting, gathering and delivering consistently over time the required data in their respective field of activities. The forthcoming national forest inventory in DRC is a typical activity that should receive such institutional support since it will bring a valuable contribution for a more efficient forest cover monitoring.

Besides, satellite Earth observation is considered as the most effective approach for a nationwide scale forest monitoring in DRC. However, the country still lacks the necessary human and logistic resources to achieve such monitoring on a regular basis. This argues in favour of strengthening the collaboration between congolese national institutions and other entities – such as OSFAC ('Observatoire Satellital des Forêts d'Afrique Centrale') and OFAC ('Observatoire des Forêts d'Afrique Centrale') – that can provide capacity building as well as contribute to the acquisition and analysis of satellite images for forest monitoring. In addition, initiatives such as REDD+ promoting the free access to satellite imagery for developing countries have to be reinforced and actively supported by national decision makers.

Human-induced forest cover changes in DRC are mainly related to the fast growing population whose livelihood is intrinsically linked to the small holder agriculture. Indeed, the DRC human population doubled over the last thirty years and simulations conducted in the present thesis illustrated that the next thirty years will likely face similar population growth figures. The doubling of the population in 2035 will result into doubling the annual rate of forest cover loss observed between 2000 and 2005. A strategic development of the agricultural sector, particularly in rural areas, is therefore critically needed since the DRC is committed to ensure a sustainable use of its forest resources. This strategic development should promote agricultural practices that guarantee better crop yields in a sustainable manner and contribute to maintain an acceptable level of soil fertility over time. Such agricultural practices are expected to reduce the progress of the agricultural frontiers into intact forest blocks.

The DRC forests are also threatened by roads construction and rehabilitation since remote forest areas will become more accessible from surrounding major towns and big villages. The research framework developed in this thesis can allow conducting the necessary impact assessment of infrastructure building and exploring various scenarios of forest cover change that can further be used as decision making tools for land-use planning. Such impact assessment can contribute to put in place multisectoral platforms which can include — among others — national stakeholders in charge of infrastructure building, agriculture and economic development, forest management, and nature conservation to explore the possible alternatives ensuring an economically viable and a sustainable use of the forest resources.

Overall, the forest cover loss modeling carried out in the present research highlighted the need for DRC policies related to a long-term national strategy of land-use planning. To be more efficient, such a land-use planning should be designed by taking into account the regional context at the scale of Congo Basin countries as well as the international context since forest cover changes in the DRC may be driven by the increasing international demand for both meat and biofuel (Megevand *et al.*, 2013). National land-use planning should therefore ultimately lead to an economic development that sustainably uses the DRC forest resources. A way to achieve such objective could be establishing a series of goals that have to be reached over time in terms of infrastructure building, agricultural production, economic development, and forest management.

Concluding statement

The analysis carried out in this thesis contributed to the credibility and the transparency of forest cover change estimates obtained from satellite Earth observation for the DRC. The research also brought some quantitative and spatially-explicit information for assessing the impact of the major drivers of forest cover change in DRC. The information have been originally retrieved in a poor data environment and allowed exploring the predictions of forest cover loss between 2005 and 2035. These predictions have shown that the DRC is likely to remain a high forested and low deforested country if mitigation actions to reduce the forest cover loss are undertaken, although the population growth coupled with infrastructure building is likely to drastically increase the rate of forest cover change. Such predictions are critical since they can support future DRC policies for a long term and strategic land-use planning that can conciliate the economic development with a sustainable use of the forest resources.

Publications related to this thesis

- **Kibambe Lubamba, J.-P.**,Desclée, B., and Defourny, P. (in preparation) Operational Deforestation Monitoring Systems using Landsat imagery: Consolidated Forest Cover Change Estimates from Pixel-Based versus Object-Based Approaches. To be submitted to *International Journal of Applied Earth Observation and Geoinformation.*
- **Kibambe Lubamba, J.-P.**, Bogaert, P., Nackoney R., J., and Defourny, P. (to be submitted) Assessing the human population distribution in African developing countries using a double-constrained-regression model. *Journal of Land Use Science.*
- **Kibambe Lubamba, J.-P.**, Radoux, J. and Defourny P. 2013. Multimodal accessibility modeling from coarse transportation networks in Africa. *International Journal of Geographical Information Science*, 27 (5): 1005 - 1022.
- **Kibambe Lubamba, J.-P.**, Jungers, Q., Rwemalika, R. and Defourny, P. (in preparation) A spatially-explicit drivers-based model of the forest cover loss in the Democratic Republic of the Congo. To be submitted to *Environmental Modeling & Sofware.*
- **Kibambe Lubamba, J.-P.** and Defourny, P. (2010). Towards Land Use Dynamics Modeling: A Case Study of the Democratic Republic of the Congo. In: *Monitoring Forest Carbon Stocks and Fluxes in the Congo Basin*, February 2-4, Brazzaville, Republic of Congo.
- Dupain, J., Nackoney, J., **Kibambe Lubamba, J.-P.**, Bokelo Bile, D., and Williams, D. 2009. Maringa-Lopori-Wamba. In de Wasseige, C. and Devers, D. and de Marcken, P. and Eba'a Atyi, R. and Nasi, R. and Mayaux, Ph. (Eds.), The Forests of the Congo Basin - State of the forest 2008. Chapter Maringa-Lopori-Wamba, pages 329-338. Publications Office of the European Union.
- Vancutsem, C., **Kibambe Lubamba, J.-P.**, Blaes, X., Pekel, J.-F., Alarcon, M. and Defourny, P. 2007. Carte générale de la République Démocratique du Congo 1 : 2 000 000. Presses Universitaires de Louvain. ISBN: 978-2-87463-061-3.

Bibliography

Achard, F., Stibig, H.-J., Eva, H. D., Lindquist, E. J., Bouvet, A., Arino, O., and Mayaux, P. 2010. Estimating tropical deforestation from Earth observation data. Carbon Management, 1(2):271–287.

Angelsen, A. 1999. Agricultural expansion and deforestation: modelling the impact of population, market forces and property rights. Journal of Development Economics, 58(1):185–218.

Angelsen, A. 2001. Playing Games in the Forest: State-Local Conflicts of Land Appropriation. Land Economics, 77(2):285–299.

Apan, A. A. and Peterson, J. A. 1998. Probing tropical deforestation: The use of GIS and statistical analysis of georeferenced data. Applied Geography, 18(2):137–152.

Archer, E. R. M. 2004. Beyond the "climate versus grazing" impasse: using remote sensing to investigate the effects of grazing system choice on vegetation cover in the eastern Karoo. Journal of Arid Environments, 57(3):381–408.

Awusabo-Asare, K., Baldwin, W., and Engebretsen, S. 2009. Demographic Data for Development: Ghana. Technical report, Population Council.

Baker, T. J., Miller, S. N., Prager, S. D., and Legg, D. E. 2010. Disaggregating human population for improved land use management in Kenya. Journal of Land Use Science, 5(4):237–257.

Baldauf, T., Plugge, D., Rqibate, A., and Köhl, M. 2009. Case studies on measuring and assessing forest degradation - Monitoring degradation in the scope of REDD. Technical report, FAO.

Balk, D. and Yetman, G. 2004. The Global distribution of population: evaluating the gains in resolution refinement. Technical report, Center for Internation Earth Science Information Network.

Bamba, I., Yedmel, M. S., and Bogaert, J. 2010. Effets des Routes et des Villes sur la forêt dense dans la Province Orientale de la République Démocratique du Congo. European Journal of Scientific Research, 43 (3):417–429.

Barredo, J. I., Kasanko, M., McCormick, N., and Lavalle, C. 2003. Modelling dynamic spatial processes: simulation of urban future scenarios through cellular automata. Landscape and Urban Planning, 64(3):145–160.

Berry, M. W., Hazen, B. C., MacIntyre, R. L., and Flamm, R. O. 1996. Lucas: a system for modeling land-use change. Computational Science Engineering, IEEE, 3(1):24–35.

Beuchle, R., Eva, H. D., Stibig, H.-J., Bodart, C., Brink, A., Mayaux, P., Johansson, D., Achard, F., and Belward, A. 2011. A satellite data set for tropical forest area change assessment. International Journal of Remote Sensing, 32:22:7009–7031.

Birch, C. P. D., Oom, S. P., and Beecham, J. A. 2007. Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. Ecological Modelling, 206(3):347–359.

Bithell, M. and Brasington, J. 2009. Coupling agent-based models of subsistence farming with individual-based forest models and dynamic models of water distribution. Environmental Modelling & Software, 24(2):173 – 190.

Blaes, X. 2009. Spécifications de contenu du Référentiel Géographique Commun - RDC. Technical report, United Nations Development Program - DRC.

Blaschke, T. 2010. Object based image analysis for remote sensing. Journal of Photogrammetry and Remote Sensing, 65:2–16.

Blaschke, T. and Strobl, J. 2001. What's wrong with pixels? Some recent developments interfacing remote sensing and GIS. In *GeoBIT/GIS*.

Bodart, C., Eva, H., Beuchle, R., Rasi, R., Simonetti, D., Stibig, H.-J., Brink, B., Lindquist, E., and Achard, F. 2011. Pre-processing of a sample of multi-scene and multi-date Landsat imagery used to monitor forest cover changes over the tropics. ISPRS Journal of Photogrammetry and Remote Sensing, 66(5):555–563.

Bogaert, J., Issouf Bamba, K., Koffi, J., Sibomana, S., Kabulu Djibu, J.-P., Champluvier, D., Robbrecht, E., De Cannière, C., , and Visser, M. N. 2008. *Patterns and Processes in Forest Landscapes - Multiple Use and Sustainable Management*, chapter Fragmentation of Forest Landscapes in Central Africa: Causes, Consequences and Management, pages 67–87. Springer.

Bousquet, F., Bakam, I., Proton, H., and Le Page, C. 1998. Cormas : common-pool resources and multi-agent systems. Lecture Notes in Artificial Intelligence, 1416:826–837.

Brady, M., de Wasseige, C., Altstatt, A., Davies, D., Mayaux, P., and Tadoum, M. 2010. Monitoring forest carbon stocks and fluxes in the Congo basin. In *Monitoring forest carbon stocks and fluxes in the Congo basin*.

Brainard, J. 1999. Integrating geographical information systems into travel cost analysis and benefit transfer. International Journal of Geographical Information Science, 13(3):227–246.

Brainard, J. S., Lovett, A. A., and Bateman, I. J. 1997. Using isochrone surfaces in travel-cost models. Journal of Transport Geography, 5(2):117–126.

Briggs, D. J., Gulliver, J., Fecht, D., and Vienneau, D. 2007. Dasymetric modelling of small-area population distribution using land cover and light emission data. Remote Sensing of Environment, 108:451–466.

Broich, M., Hansen, M. C., Potapov, P., Adusei, B., Lindquist, E., and Stehman, S. V. 2011. Time-series analysis of multi-resolution optical imagery for quantifying forest cover loss in Sumatra and Kalimantan, Indonesia. International Journal of Applied Earth Observation and Geoinformation, 13:277–291.

Brown, D. G., Pijanowski, B., and Duh, J. D. 2000. Modeling the relationships between land use and land cover on private lands in the Upper Midwest, USA. Journal of Environmental Management, 59(4):247–263.

Brown, D. G., Goovaerts, P., Burnicki, A., and Li, M.-Y. 2002. Stochastic Simulation of Land-Cover Change Using Geostatistics and Generalized Additive Models. Photogrammetric Engineering & Remote Sensing, 68(10):1051–1061.

Brown, D. G., Walker, R., Manson, S., and Seto, K. 2004. Modeling Land Use and Land Cover Change. In Gutman, G., Janetos, A. C., Justice, C. O., Moran, E. F., Mustard, J. F., Rindfuss, R. R., Skole, D., Turner, B. L., Cochrane, M. A., and Meer, F. D., editors, *Land Change Science*, volume 6 of *Remote Sensing and Digital Image Processing*, pages 395–409. Springer Netherlands.

Cabral, A. I. R., Vasconcelos, M. J., Oom, D., and Sardinha, R. 2011. Spatial dynamics and quantification of deforestation in the central-plateau woodlands of Angola (1990 à 2009). Applied Geography, 31(3):1185–1193.

Castella, J.-C., Manh, P. H., Kam, S. P., Villano, L., and Tronche, N. R. 2005. Analysis of village accessibility and its impact on land use dynamics in a mountainous province of northern Vietnam. Applied Geography, 25(4):308–326.

Castle, C. J. and Crooks, A. T. 2006. Principles and Concepts of Agent-Based Modelling for Developing Geospatial Simulations. In *University Colledge London Working Papers Series - Paper 110*. Centre for Advanced Spatial Analysis.

Cervero, R. and Radisch, C. 1996. Travel choices in pedestrian versus automobile oriented neighborhoods. Transport Policy, 3(3):127–141.

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

CIA. 2010. *The World Factbook*. US Government Printing Office.

Clark, S., Bolt, K., and Campbell, A. 2008. Protected areas: an effective tool to reduce emissions from deforestation and forest degradation in developing countries? Technical report, UNEP-WCMC.

Colom, A. 2006. Aspects socio-économiques de l'utilisation et de la gestion des ressources naturelles dans le paysage Salonga-Lukenie-Sankuru: un guide pour la conservation et l'amélioration des conditions de vie. Technical report, World Wildlife Foundation - DRC.

Congalton, R. G. 1997. Exploring and Evaluating the Consequences of Vector-to-Raster and Raster-to-Vector Conversion. Photogrammetic Engineering & Remote Sensing, 63(4):425–434.

Cracknell, A. P. 1998. Synergy in remote sensing - what's in a pixel? International Journal of Remote Sensing, 19:2025–2047.

Crampton, J. W. 2004. GIS and Geographic Governance: Reconstructing the Chloropleth Map. Cartographica, 39:41–53.

Curtin, K. M., Nicoara, G., and Arifin, R. R. 2007. A Comprehensive Process for Linear Referencing. URISA Journal, 19(2):23–32.

Czaplewski, R. 2003. Can a sample of Landsat sensor scenes reliably estimate the global extent of tropical deforestation? International Journal of Remote Sensing, 24(6):1409–1412.

Davies, D., Ilavajhala, S., Wong, M., and Justice, C. 2009. Fire Information for Resource Management System: Archiving and Distributing MODIS Active Fire Data. IEEE Transactions on Geoscience and Remote Sensing, 47:72–79.

de Jong, T. and van Eck, J. R. 1996. Location profile-based measures as an improvement on accessibility modelling in GIS. Computers, Environment and Urban Systems, 20(3):181–190.

De Saint Moulin, L. 2006. Analyse par territoire et ville des résultats de l'enrôlement des élécteurs et du référendum sur le projet de constitution. Congo-Afrique, Numéro Spécial:9–29.

de Sousa, L., Nery, F., Sousa, R., and Matos, J. 2006. Assessing the accuracy of hexagonal versus squared tilled grids in preserving DEM surface flow directions. In Caetano, M. and eds. M. Painho, editors, *Proceedings of the 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences*.

Debroux, L., Hart, T., Kaimowitz, D., Karsenty, A., and Topa, G. E., editors. 2007. *Forest in post-conflict Democratic Republic of Congo: Analysis of a Priority Agenda*. Center for International Forestry Research.

Defourny, P., Bicheron, P., Carsten, B., Bontemps, S., Van Bogaert, E., Vancutsem, C., Pekel, J.-F., Huc, M., Henry, C. C., Ranera, F., Achard, F., Di Gregorio, A., Herold, M., Leroy, M., and Arino, O. 2009. The first 300 m global land cover map for 2005 using ENVISAT MERIS time series : a product of the GlobCover system. In *33rd International Symposium on Remote Sensing of Environment*.

Defourny, P., Delhage, C., and Kibambe Lubamba, J.-P. 2011. Analyse quantitative des causes de la déforestation et de la dégradation des forêts en République Démocratique du Congo. Technical report, Earth and Life Institue Environment - Université catholique de Louvain.

DeFries, R. S., Achard, F., Brown, S., Herold, M., Murdiyarso, D., Schlamadinger, B., and de Souza, C. 2007. Earth observations for estimating greenhouse emissions from deforestation in developing countries. Environmental Science and Policy, 10:385–394.

Deichmann, U., Balk, D., and Yetman, G. 2001. Transforming population data for interdisciplinary usages: from census to grid. Technical report, Center for International Earth Science Information Network.

Dendoncker, N., Bogaert, P., and Rounsevell, M. 2006. A statistical method to downscale aggregated land use data and scenarios. Journal of Land Use Science, 1(2-4):63–82.

DePencier, D. 2008. Measuring Multi-Modal Travel Costs in the Vicinity of Limited-Access Transportation Networks. Technical report, University of Pennsylvania.

Desclée, B., Bogaert, P., and Defourny, P. 2006. Forest change detection by statistical object-based method. Remote Sensing of Environment, 102:1–11.

Di Gregorio, A. 2005. *Land Cover Classification System. Classification concepts and user manual. Software version 2*. FAO.

Dobson, J. E., Bright, E. A., Coleman, P. R., Durfee, R. C., and Worley, B. A. 2000. LandScan: A global population database for estimating population at risk. Photogrammetric Engineering & Remote Sensing, 66 (7):849–857.

Dooley, J. F. 2005. An inventory and comparison of globally consistent geospatial databases and libraries. Technical report, UN Food and Agriculture Organization.

Duveiller, G., Defourny, P., Desclée, B., and Mayaux, P. 2008. Deforestation in Central Africa: Estimates at regional, national and landscape levels by advanced processing of systematically-distributed Landsat extracts. Remote Sensing of Environment, 112(5):1969–1981.

Eastman, J. R. 1999. *IDRISI 32: IDRISI for workstations*. Worcester: Clark University.

Eastman, J. R. 2009. *Idrisi Taiga, guide to GIS and remote processing*, chapter The Land Change Modeler for Ecological Sustainability, pages 235–342. Worcester: Clark University.

Eastman, J. R., Solorzano, L. A., and van Fossen, M. E. 2005. *GIS, Spatial Analysis, and Modeling*, chapter Transition Potential Modeling for land-cover change, pages 357–385. ESRI Publisher Press.

Echeverria, C., Coomes, D., Salas, J., Rey-Benayas, J. M., Lara, A., and Newton, A. 2006. Rapid deforestation and fragmentation of Chilean Temperate Forests. Biological Conservation, 130:481–494.

Eckert, S., Ratsimba, H. R., Rakotondrasoa, L. O., Rajoelison, L. G., and Ehrensperger, A. 2011. Deforestation and forest degradation monitoring and assessment of biomass and carbon stock of lowland rainforest in the Analanjirofo region, Madagascar. Forest Ecology and Management, 262:11:1996–2007.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroehl, H. W., and Davis, E. R. 1997. Mapping city lights with nighttime data from the DMSP operational linescan system. Photogrammetric Engineering & Remote Sensing, 63:727–734.

Ernst, C., Mayaux, P., Verhegghen, A., Bodart, C., Musampa, C., and Defourny, P. 2013. National forest cover change in Congo Basin: deforestation, reforestation, degradation and regeneration for the years 1990, 2000 and 2005. Global Change Biology, 19(4):1173–1187.

FAO. 2001. Global forest resources assessment 2000. Technical report, Food and Agriculture Organization of the United Nations.

FAO. 2006. Global forest resources assessment 2005. Technical report, Food and Agriculture Organization of the United Nations.

FAO. 2010. Global Forest Resources Assessment 2010 - Terms and Definitions. Technical report, Food and Agriculture Organization of the United Nations.

Flowerdew, R. and Green, M. 1992. Developments in areal interpolation methods and GIS. The Annals of Regional Science, 26(1):67–78.

Gallopin, G. C. 2006. Linkages between vulnerability, resilience, and adaptive capacity. Global Environmental Change, 16(3):293–303.

Geurs, K. T. and van Eck, J. R. 2001. Accessibility measures: review and applications. Technical report, Urban Research Centre.

Geurs, K. T. and Van Wee, B. 2004. Accessibility evaluation of land-use and transport strategies: review and research directions. Journal of Transport Geography, 12:127–140.

Geweke, J. 1986. Exact inference in the inequality constrained normal linear regression model. Journal of Applied Econometrics, 1:127–141.

Gil Pontius Jr., R. G., , Schneider, L. C., Schneider, L. C., Gil Pontius Jr., R. G., , Schneider, L. C., Schneider, L. C., and Gil Pontius Jr., R. G. 2001. Land-cover change model validation by an ROC method for the Ipswich watershed, Massachusetts, USA. Agriculture, Ecosystems and Environment, 85(1-3):239–248.

Gleyze, J.-F. and Rousseaux, F. 2003. Impact of relief accuracy on flood simulations and road network vulnerability analysis. In *13th European Colloquium on Theoretical and Quantitative (ecQTG'03)*, page 14 pages. Lucca (Italy).

Goodchild, M. F., Anselin, L., and Deichmann, U. 1993. A Framework for the Areal Interpolation of Socioeconomic Data. Environment and Planning A, 25:383–397.

Gun Wang, H., Federico Barra, A., Krishnamani, R., and Nogales, A. 2009. Déterminer les priorités des investissements dans l'infrastructure: une approche spatiale. Technical report, The World Bank.

Gutiérrez, J. and Urbano, P. 1996. Accessibility in the European Union: the impact of the trans-European road network. Journal of Transport Geography, 4(1):15–25.

Gutiérrez, J., Gonzalez, R., and Gomez, G. 1996. The European high-speed train network : Predicted effects on accessibility patterns. Journal of Transport Geography, 4(4):227–238.

Haberl, H., Batterbury, S., and Moran, E. 2001a. Using and shaping the land: a long-term perspective. Land Use Policy, 18(1):1–8.

Haberl, H., Erb, K.-H., Krausmann, F., Loibl, W., Schulz, N., and Weisz, H. 2001b. Changes in ecosystem processes induced by land use: Human appropriation of aboveground NPP and its influence on standing crop in Austria. Global Biogeochemical Cycles, 15(4):929–942.

Haeusler, T., Amougou, J., Gomez, S., Siwe, R., Ramminger, G., and Seifert-Granzin, J. 2010. Development and Implementation of GSE FM REDD Pilot Projects in the Congo Region / REDD project in Cameroon. In *Monitoring forest carbon stocks and fluxes in the Congo basin*.

Hansen, M. C., Roy, D. P., Lindquist, E., Adusei, B., Justice, C. O., and Altstatt, A. 2008a. A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo basin. Remote Sensing of Environment, 112:2495–2513.

Hansen, M. C., Stehman, S. V., Potapov, P. V., Loveland, T. R., Townshend, J. R. G., DeFries, R. S., Pittman, K. W., Arunarwati, B., Stolle, F., Steininger, M. K., Carroll, M., and DiMiceli, C. 2008b. Humid tropical forest clearing from 2000 to 2005 quantified by using multitemporal and multiresolution remotely sensed data. Proceedings of the National Academy of Sciences.

Hay, S., Noor, A., Nelson, A., and Tatem, A. 2005. The accuracy of human population maps for public health application. Tropical Medicine and International Health, 10:1–14.

Herderschee, J., Kaiser, K.-A., and Samba Mukoko, D. 2012. *Résilience d'un Géant Africain - Accélérer la croissance et promouvoir l'emploi en République Démocratique du Congo*, volume 1, chapter Synthèse, Contexte Historique et Macroéconomique, pages 1–352. Médiaspaul.

Herold, M., Latham, J. S., Di Gregorio, A., and Schmullius, C. C. 2006. Evolving standards in land cover characterization. Journal of Land Use Science, 1:2-4:157–168.

Herold, M., Mayaux, P., Woodcock, C. E., Baccini, A., and Schmullius, C. C. 2008. Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. Remote Sensing of Environment, 112:2538–2556.

Hibbard, K., Janetos, A., van Vuuren, D. P., Pongratz, J., Rose, S. K., Betts, R., Herold, M., and Feddema, J. J. 2010. Research priorities in land use and land-cover change for the Earth system and integrated assessment modelling. International Journal of Climatology, 30(13):2118–2128.

Hinton, G. E. and Salakhutdinov, R. R. 2006. Reducing the Dimensionality of Data with Neural Networks. Science, 313(5786):504–507.

Hsiao, C. 2006. Panel Data Analysis - Advantages and Challenges. Technical report, University of Southern California and Wang Yanan Institute for Studies in Economics.

Iaych, K. 2010. Word road statistics 2010 - Data 2003 to 2008. Technical report, Geneva.

Janetos, A. C. 2004. Research Directions in Land-Cover and Land-Use Change. In Gutman, G., Janetos, A. C., Justice, C. O., Moran, E. F., Mustard, J. F., Rindfuss, R. R., Skole, D., Turner, B. L., Cochrane, M. A., and Meer, F. D., editors, *Land Change Science*, volume 6 of *Remote Sensing and Digital Image Processing*, pages 449–457. Springer Netherlands.

Jepson, W. and Millington, A. 2008. *Land-Change Science in the Tropics*, chapter The Changing Countryside, pages 1–10. Springer.

Juliao, R. P. 1999. Accessibility and GIS. In *39th Congress of ERSA / Dublin, 23-27 August*. ERSA.

Jung, M., Henkel, K., Herold, M., and Churkina, G. 2006. Exploiting synergies of global land cover products for carbon cycle modeling. Remote Sensing of Environment, 101:534–553.

Kamusoko, C., Aniya, M., Adi, B., and Manjoro, M. 2009. Rural sustainability under threat in Zimbabwe - Simulation of future land use/cover changes in the Bindura district based on the Markov-cellular automata model. Applied Geography, 29:435–447.

Karsenty, A. and Ongolo, S. 2012. Can "fragile states" decide to reduce their deforestation? The inappropriate use of the theory of incentives with respect to the REDD mechanism. Forest Policy and Economics, 18:28–45.

Khoi, D. D. and Murayama, Y. 2011. *Spatial analysis and modeling in geographical transformation process*, chapter Modeling deforestation using a neural network-Markov model, pages 169–190. Springer.

Lambin, E. 1999. Monitoring forest degradation in tropical regions by remote sensing: some methodological issues. Global Ecology and Biogeography, 8:191–198.

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., Rounsevell, M. D. A., and Geist, H. J. 2000. Are agricultural land-use models able to predict changes in land-use intensity? Agriculture, Ecosystems & Environment, 82(1-3):321–331.

Lambin, E. F., Geist, H. J., and Lepers, E. 2003. Dynamics of land-use and land-cover change in tropical regions. Annual Review of Environment and Resources, 28(1):205–241.

Lambin, E. F., Geist, H., and Rindfuss, R. R. 2006. *Land-Use and Land-Cover Change - Local Processes and Global Impacts*, chapter Introduction: Local Process with Global Impacts, pages 1–8. Springer.

Langford, M. 2006. Obtaining population estimates in non-census reporting zones: An evaluation of the 3-class dasymetric method. Computers, Environment and Urban Systems, 30(2):161–180.

Langford, M., Maguire, D. J., and Unwin, D. J. 1991. *Handling Geographical Information: Methodology and Potential Applications*, chapter The areal interpolation problem: Estimating population using remote sensing in a GIS framework, pages 55–77. London: Longman Press.

Laporte, N. T., Stabach, J. A., Grosch, R., Lin, T. S., and Goetz, S. J. 2007. Expansion of Industrial Logging in Central Africa. Science, 316(5830):1451.

Laris, P. 2002. Burning the Seasonal Mosaic: Preventative Burning Strategies in the Wooded Savanna of Southern Mali. Human Ecology, 30(2):155–186.

Latham, J. 2001. Africover eastern Africa. LUCC Newsletter, 7:15–16.

Linard, C., Gilbert, M., Snow, R. W., Noor, A. M., and Tatem, A. J. 2012. Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. PLoS ONE, 7:e31743.

Lindquist, E. J., Hansen, M. C., Roy, D. P., and Justice, C. O. 2008. The suitability of decadal image data sets for mapping tropical forest cover change in the Democratic Republic of Congo: implications for the global land survey. International Journal of Remote Sensing, 29:24:7269–7275.

Liu, J., Ouyang, Z., Taylor, W. W., Groop, R. E., Tan, Y., and Zhang, H. 1999. Framework for Evaluating the Effects of Human Factors on Wildlife Habitat: the Case of Giant Pandas. Conservation Biology, 13(6):1360–1370.

Liu, J., An, L., Batie, S. S., Groop, R. E., Liang, Z., Linderman, M. A., Mertig, A. G., Ouyang, Z., and Qi, J. 2004. Human Impacts on land Cover and Panda Habitat in Wolong Nature Reserve. In Fox, J., Rindfuss, R., Walsh, S., and

Mishra, V., editors, *People and the Environment*, pages 241–263. Springer US.

Liu, J., Dietz, T., Carpenter, S. R., Alberti, M., Folke, C., Moran, E., Pell, A. N., Deadman, P., Kratz, T., Lubchenco, J., Ostrom, E., Ouyang, Z., Provencher, W., Redman, C. L., Schneider, S. H., and Taylor, W. W. 2007. Complexity of Coupled Human and Natural Systems. Science, 317(5844):1513–1516.

Liu, S. and Zhu, X. 2004a. Accessibility Analyst: an integrated GIS tool for accessibility analysis in urban transportation planning. Environment and Planning B: Planning and Design, 31(1):105–124.

Liu, S. and Zhu, X. 2004b. An integrated GIS approach to accessibility analysis. Transactions in GIS, 8 (1):45–62.

Liu, X. H., Kyriakidis, P. C., and Goodchild, M. F. 2008. Population-density estimation using regression and area-to-point residual kriging. International Journal of Geographical Information Science, 22(4):431–447.

Lo, C. P. 2008. Population estimation using geographically weighted regression. GIScience & Remote Sensing, 45 (2):131–148.

Lu, H. and Liu, G. 2013. Distributed land use modeling and sensitivity analysis for REDD+. Land Use Policy, 33(0):54 – 60.

Ludeke, A. K., Maggio, R. C., and Reid, L. M. 1990. An analysis of Anthropogenic Deforestation Using Logistic Regression and GIS. Journal of Environmental Management, 31:247–259.

Makana, J.-R. and Thomas, S. C. 2006. Impacts of Selective Logging and Agricultural Clearing on Forest Structure, Floristic Composition and Diversity, and Timber Tree Regeneration in the Ituri Forest, Democratic Republic of Congo. Biodiversity & Conservation, 15(4):1375–1397.

Martin, D., Wrigley, H., Barnett, S., and Roderick, P. 2002. Increasing the sophistication of access measurement in a rural healthcare study. Health & Place, 8(1):3–13.

Mas, J. F., Puig, H., Palacio, J. L., and Sosa-López, A. 2004. Modelling deforestation using GIS and artificial neural networks. Environmental Modelling & Software, 19(5):461–471.

Mayaux, P., Holmgren, P., Achard, F., Eva, H., Stibig, H.-J., and Branthomme, A. 2005. Tropical forest cover change in the 1990s and options for future monitoring. Philosophical Transactions for the Royal Society B, 360:373–384.

McCallum, I., Obersteiner, M., Nilsson, S., and Shvidenko, A. 2006. A spatial comparison of four satellite derived 1 km global land cover datasets. International Journal of Applied Earth Observation and Geoinformation, 8:246–255.

McGrail, M. R. and Humphreys, J. S. 2009. Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method. Applied Geography, 29(4):533–541.

McRoberts, R. E., Cohen, W. B., Naesset, E., Stehman, S. V., and Tomppo, E. O. 2010. Using remotely sensed data to construct and assess forest attribute maps and related spatial products. Scandinavian Journal of Forest Research, 25:4:340–367.

Megevand, C., Mosnier, A., Hourticq, J., Sanders, K., Doetinchem, N., and Streck, C. 2013. *Deforestation trends in the Congo basin - Reconciling economic growth and forest protection*. The World Bank, Washington DC, United States of America.

Mennis, J. 2009. Dasymetric Mapping for Estimating Population in Small Areas. Geography Compass, 3(2):727–745.

Mertens, B. and Lambin, E. F. 2000. Land-Cover-Change Trajectories in Southern Cameroon. Annals of the Association of American Geographers, 90(3):467–494.

Moisi, J., Gatakaa, H., Noor, A., Williams, T., Bauni, E., Tsofa, B., Levine, O., and Scott, J. A. 2010. Geographic access to care is not a determinant of child mortality in a rural Kenyan setting with high health facility density. BMC Public Health, 10:142.

Monmonier, M. 1996. How to lie with maps. chapter Data maps: Making nonsense of the census, pages 123–146. Chicago: The University of Chicago Press.

Mooney, H., Larigauderie, A., Cesario, M., Elmquist, T., Hoegh-Guldberg, O., Lavorel, S., Mace, G. M., Palmer, M., Scholes, R., and Yahara, T. 2009. Biodiversity, climate change, and ecosystem services. Current Opinion in Environmental Sustainability, 1(1):46–54.

Moran, E. F., Skole, D. L., and Turner, B. L. 2004. The Development of the International Land-Use and Land-Cover Change (LUCC) Research Program and Its Links to NASA's Land-Cover and Land-Use Change (LCLUC) Initiative. In Gutman, G., Janetos, A. C., Justice, C. O., Moran, E. F., Mustard, J. F., Rindfuss, R. R., Skole, D., Turner, B. L., Cochrane, M. A., and Meer, F. D., editors, *Land Change Science*, volume 6 of *Remote Sensing and Digital Image Processing*, pages 1–15. Springer Netherlands.

Moxey, A. and Allanson, P. 1994. Areal interpolation of spatially extensive variables: A comparison of alternative techniques. International Journal of Geographical Information Systems, 8:479–487.

Nackoney, J. and Williams, D. 2012. Conservation prioritization and planning with limited wildlife data in a Congo Basin forest landscape: assessing human threats and vulnerability to land use change. Journal of Conservation Planning, 8:25–44.

Nackoney, J., Rybock, D., Dupain, J., and Facheux, C. 2013. Coupling participatory mapping and GIS to inform village-level agricultural zoning in the Democratic Republic of the Congo. Landscape and Urban Planning, 110(0):164–174.

Nelson, A. 2000. Accessibility, transport and travel time information. Technical report, CIAT Hillsides.

Nelson, A. 2008. Travel time to major cities: A global map of Accessibility. Technical report, Global Environment Monitoring Unit - Joint Research Centre of the European Commission, Ispra Italy.

Nelson, G. C. and Geoghegan, J. 2002. Deforestation and land use change: sparse data environments. Agricultural Economics, 27(3):201–216.

Ngondo, P. S., De Sain Moulin, L., and Tambashe, O. B. 1992. Perspectives démographiques du Zaïre 1984-1999 & Population d'âge électoral en 1993-1994. Centre d'Etudes pour l'Action Sociale.

O'Sullivan, D., Morrison, A., and Shearer, J. 2000. Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach. International Journal of Geographical Information Science, 14(1):85–104.

Perry, B. and Gesler, W. 2000. Physical access to primary health care in Andean Bolivia. Social Science & Medicine, 50(9):1177–1188.

Peterson, G. D., Cumming, G. S., and Carpenter, S. R. 2003. Scenario Planning: a Tool for Conservation in an Uncertain World. Conservation Biology, 17(2):358–366.

Pflugmacher, D., Krankina, O. N., Cohen, W. B., Friedl, M. A., Sulla-Menashe, D., Kennedy, R. E., Nelson, P., Loboda, T. V., Kuemmerle, T., Dyukarev, E., Elsakov, V., and Kharuk, V. I. 2011. Comparison and assessment of coarse resolution land cover maps for Northern Eurasia. Remote Sensing of Environment, 115(12):3539–3553.

Piao, S., Friedlingstein, P., Ciais, P., de Noblet-Ducoudré, N., Labat, D., and Zaehle, S. 2007. Changes in climate and land use have a larger direct impact than rising CO2 on global river runoff trends. Proceedings of the National Academy of Sciences, 104(39):15242–15247.

Pijanowski, B. C., Brown, D. G., Shellito, B. A., and Manik, G. A. 2002. Using neural networks and GIS to forecast land use changes: a Land Transformation Model. Computers, Environment and Urban Systems, 26(6):553–575.

Polhill, J. G., Gotts, N. M., and Law, A. N. R. 2001. Imitative versus nonimitative strategies in a land-use simulation. Cybernetics and Systems, 32(1-2):285–307.

Pongratz, J., Reick, C., Raddatz, T., and Claussen, M. 2008. A reconstruction of global agricultural areas and land cover for the last millennium. Global Biogeochemical Cycles, 22(3):1–16.

Pontius Jr, R. G., Cornell, J. D., and Hall, C. A. S. 2001. Modeling the spatial pattern of land-use change with GEOMOD2: application and validation for Costa Rica. Agriculture, Ecosystems and Environment, 85(1):191–203.

Potapov, P. V., Turubanova, S. A., Hansen, M. C., Adusei, B., Broich, M., Altstatt, A., Mane, L., and Justice, C. O. 2012. Quantifying forest cover loss in Democratic Republic of the Congo, 2000-2010, with Landsat ETM+ data. Remote Sensing of Environment, 122(0):106–116.

Putz, F. 2010. The Importance of Defining 'Forest': Tropical Forest Degradation, Deforestation, Long-term Phase Shifts, and Further Transitions. Biotropica, 42(1):10–20.

Radoux, J., Defourny, P., and Bogaert, P. 2008. Comparison of pixel- and object-based sampling strategies for thematic accuracy assessment. In G.J.Hay, T. B. and editors D. Marceau, editors, *GEOBIA 2008-Pixels, Objects, Intelligence. GEOgraphic Object Based Image Analysis for the 21st Century.*, volume XXXVIII-4/. University of Calgary, ISPRS.

Radoux, J., Bogaert, P., Fasbender, D., and Defourny, P. 2011. Thematic accuracy assessment of geographic object-based image classification. International Journal of Geographical Information Science, 25(6):895–911.

Rasi, R., Bodart, C., Stibig, H.-J., Eva, H., Beuchle, R., Carboni, S., Simonetti, D., and Achard, F. 2011. An automated approach for segmenting and classifying a large sample of multi-date Landsat imagery for pan-tropical forest monitoring. Remote Sensing of Environment, 115:3659–3669.

Reibel, M. and Agrawal, A. 2007. Areal Interpolation of Population Counts Using Pre-classified Land Cover Data. Population Research and Policy Review, 26(5):619–633.

Reibel, M. and Bufalino, M. E. 2005. Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. Environment and Planning A, 37(1):127–139.

Riebsame, W. E., Parton, W. J., Galvin, K. A., Burke, I. C., Bohren, L., Young, R., and Knop, E. 1994. Integrated Modeling of Land Use and Cover Change. BioScience, 44(5):350–356.

Rindfuss, R. R., Walsh, S. J., Turner, B. L., Fox, J., and Mishra, V. 2004. Developing a science of land change: Challenges and methodological issues. Proceedings of the National Academy of Sciences of the United States of America, 101(39):13976–13981.

Robbins, P. 2001. Tracking Invasive Land Covers in India, or Why Our Landscapes Have Never Been Modern. Annals of the Association of American Geographers, 91(4):637–659.

Rockström, J., Steffen, W., Noone, K., Person, A., Stuart, C., Lambin, E., Lenton, T., Scheffer, M., Folke, C., Schellnhuber, H., Nykvist, B., de Wit, C., Hughes, T., van der Leeuw, S., Rodhe, H., Sörlin, S., Snyder, P., Costanza, R., Svedin, U., Falkenmark, M., Karlberg, L., Corell, R., Fabry, V., Hansen, J., Walker, B., Liverman, D., Richardson, K., Crutzen, P., and Fo., J. 2009. A safe operating space for humanity. Nature, 461:472–475.

Rounsevell, M. D. A., Pedroli, B., Erb, K.-H., Gramberger, M., Busck, A. G., Haberl, H., Kristensen, S. r., Kuemmerle, T., Lavorel, S., Lindner, M., Lotze-Campen, H., Metzger, M. J., Murray-Rust, D., Popp, A., Pérez-Soba, M., Reenberg, A., Vadineanu, A., Verburg, P. H., and Wolfslehner, B. 2012. Challenges for land system science. Land Use Policy, 29(4):899–910.

Salvatore, M., Pozzi, F., Ataman, E., Huddleston, B., and Bloise, M. 2005. Mapping global urban and rural population distributions. Technical report, UN Food and Agriculture Organization.

Sangermano, F., Eastman, J. R., and Zhu, H. 2010. Similarity Weighted Instance-based Learning for the Generation of Transition Potentials in Land Use Change Modeling. Transactions in GIS, 14(5):569–580.

Scheiner, J. 2010. Interrelations between travel mode choice and trip distance: trends in Germany 1976-2002. Journal of Transport Geography, 18:75–84.

Scheiner, J. and Holz-Rau, C. 2007. Travel mode choice: affected by objective or subjective determinants? Transportation, 34:487–511.

Schneider, R. R. 1995. *Government and the economy on the Amazon frontier*. World Bank environment paper ;. World Bank, Washington, DC.

Schneider, U. A., Havlík, P., Schmid, E., Valin, H., Mosnier, A., Obersteiner, M., Böttcher, H., Skalský, R., Balkovič, J., Sauer, T., and Fritz, S. 2011. Impacts of population growth, economic development, and technical change on global food production and consumption. Agricultural Systems, 104(2):204 – 215.

Schure, J., Ingram, V., and Akalakou-Mayima, C. 2011. Bois énergie en RDC: Analyse de la filière des villes de Kinshasa et de Kisangani. Technical report, Projet Makala/CIFOR.

Sertel, E., Robock, A., and Ormeci, C. 2010. Impacts of land cover data quality on regional climate simulations. International Journal of Climatology, 30(13):1942–1953.

Small, C., Pozzi, F., and Elvidge, C. D. 2005. Spatial analysis of global rural extent from DMSP-OLS night lights. Remote Sensing of Environment, 96:277–291.

Soares-Filho, B. S., Cerqueira, G. C., and Pennachin, C. L. 2002. Dinamica: a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. Ecological Modelling, 154(3):217–235.

Srinivasan, S. and Rogers, P. 2005. Travel behavior of low-income residents: studying two contrasting locations in the city of Chennai, India. Journal of Transport Geography, 13:265–274.

Stannard, C. A. and Aspinall, R. J. 2011. Meeting the challenges of modelling coupled human-environmental systems - GLP Nodal Office of Integration and Modelling. Procedia Environmental Sciences, 6(0):194–198.

Stibig, H.-J., Belward, A. S., Roy, P. S., Rosalina-Wasrin, U., Agrawal, S., Joshi, P. K., Beuchle, R., Fritz, S., Mubareka, S., and Giri, C. 2007. A land-cover map for South and Southeast Asia derived from SPOT-VEGETATION data. Journal of Biogeography, 34(4):625–637.

Sutton, P., Roberts, D., Elvidge, C. D., and Meij, H. 1997. A comparison of nighttime satellite imagery and population density for the continental United States. Photogrammetric Engineering & Remote Sensing, 63 (11):1303–1313.

Swetnam, R. D., Fisher, B., Mbilinyi, B. P., Munishi, P. K. T., Willcock, S., Ricketts, T., Mwakalila, S., Balmford, A., Burgess, N. D., Marshall, A. R., and Lewis, S. L. 2011. Mapping socio-economic scenarios of land cover change: A GIS method to enable ecosystem service modelling. Journal of Environmental Management, 92(3):563–574.

Tadoum, M., Seya Makonga, V. K., Boundzanga, G. C., Bouyer, O., Hamel, O., Creighton, G. K., Ernst, C., Verhegghen, V., Mayaux, P., Hansen, M., and Defourny, P. 2012. The Forests of the Congo Basin - State of the Forest 2010. chapter International negociations on the future climate regime beyond 2012: Achievements from Copenhagen to Cancun and benefits to the forests of the Congo basin, pages 157–170. Office of the European Union.

Tanser, F., Gijsbertsen, B., and Herbst, K. 2006. Modelling and understanding primary health care accessibility and utilization in rural South Africa: An exploration using a geographical information system. Social Science & Medicine, 63(3):691–705.

Tatem, A., Campiz, N., Gething, P., Snow, R., and Linard, C. 2011. The effects of spatial population dataset choice on estimates of population at risk of disease. Population Health Metrics, 9(1):4.

Tatem, A. J., Noor, A. M., Von Hagen, C., and Hay, S. I. 2007. High resolution population maps for low income nations: combining land cover and census in East Africa. PLoS ONE, 2 (12).

Tian, Y., Yue, T., Zhu, L., and Clinton, N. 2005. Modeling population density using land cover data. Ecological Modelling, 189 (1-2):72–88.

Tollens, E. 2003. Current situation of food security in the D.R. Congo: Diagnostic and perspectives. Technical report, Katholieke Universiteit Leuven.

Tollens, E. 2010. Potential Impacts of Agriculture Development on the Forest Cover in the Congo Basin. Technical report, Katholieke University of Leuven (K.U.Leuven).

Torrens, P. M. and O'Sullivan, D. 2000. Cities, cells, and cellular automata: Developing a research agenda for urban geocomputation. In *Proceedings of the Fifth Annual Conference on GeoComputation*.

Townshend, J. R., Justice, C. O., Skole, D. L., Belward, A., Janetos, A., Gunawan, I., Goldammer, J. G., and Lee, B. 2004. Meeting the Goals of GOFC. In Gutman, G., Janetos, A. C., Justice, C. O., Moran, E. F., Mustard, J. F., Rindfuss, R. R., Skole, D., Turner, B. L., Cochrane, M. A., and Meer, F. D., editors, *Land Change Science*, volume 6 of *Remote Sensing and Digital Image Processing*, pages 31–55. Springer Netherlands.

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

Turner, B. L., Moran, E., and Rindfuss, R. 2004. Integrated Land-Change Science and Its Relevance to the Human Sciences. In Gutman, G., Janetos, A. C., Justice, C. O., Moran, E. F., Mustard, J. F., Rindfuss, R. R., Skole, D., Turner, B. L., Cochrane, M. A., and Meer, F. D., editors, *Land Change Science*, volume 6 of *Remote Sensing and Digital Image Processing*, pages 431–447. Springer Netherlands.

Turner, B. L., Lambin, E. F., and Reenberg, A. 2007. The emergence of land change science for global environmental change and sustainability. Proceedings of the National Academy of Sciences, 104(52):20666–20671.

UNDP. 2008. *World population prospects, 2008 revision*. United Nations Population Division.

van Wagtendonk, J. W. and James, M. B. 1980. Travel time variation on backcountry trails. Journal of Leisure Research, 12:99–106.

Vancutsem, C., Kibambe Lubamba, J.-P., Blaes, X., Pekel, J.-F., Alarcon, M., and Defourny, P. 2007. *Carte générale de la République Démocratique du Congo 1 : 2 000 000*. Presses Universitaires de Louvain.

Vancutsem, C., Pekel, J.-F., Evrard, C., Malaisse, F., and Defourny, P. 2008. Mapping and characterizing the vegetation types of the Democratic Republic of Congo using SPOT VEGETATION time series. International Journal Applied Earth Observation and Geoinformation, 11 (1):62–76.

Vande Weghe, J. and Devers, D. 2006. *State of the forests 2006*, chapter Conservation, pages 17–28. Publication Office of the European Union.

Vandenbulcke, G., Steenberghen, T., and Thomas, I. 2009. Mapping accessibility in Belgium: a tool for land-use and transport planning? Journal of Transport Geography, 17:39–53.

Verburg, P. H., de Koning, G. H. J., Kok, K., Veldkamp, A., and Bouma, J. 1999. A spatial explicit allocation procedure for modelling the pattern of land use change based upon actual land use. Ecological Modelling, 116(1):45–61.

Verburg, P. H., Veldkamp, W. S. A., Espaldon, R. L. V., and Mastura, S. S. A. 2002. Modeling the Spatial Dynamics of Regional Land Use: The CLUE-S Model. Environmental Management, 30 (3):391–405.

Verburg, P. H., de Nijs, T. C. M., van Eck, J. R., Visser, H., and de Jong, K. 2004a. A method to analyse neighbourhood characteristics of land use patterns. Computers, Environment and Urban Systems, 28:667–690.

Verburg, P. H., Overmars, K. P., and Witte, N. 2004b. Accessibility and land-use patterns at the forest fringe in the northeastern part of the Philippines. Geographical Journal, 170(3):238–255.

Verburg, P. H., Kok, K., Pontius, R. G., and Veldkamp, A. 2006. Modeling Land-Use and Land-Cover Change. In Lambin, E. F. and Geist, H., editors, *Land-Use and Land-Cover Change*, Global Change — The IGBP Series (closed), pages 117–135. Springer Berlin Heidelberg.

Verhegghen, A., Defourny, P., Ernst, C., and Beuchle, R. 2010. Automated land cover mapping and independent change detection in tropical forest using multi-temporal high resolution data set. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2010).*

Verhegghen, A., Mayaux, P., de Wasseige, C., and Defourny, P. 2012. Mapping Congo Basin vegetation types from 300 m and 1 km multi-sensor time series for carbon stocks and forest areas estimation. Biogeosciences, 9(12):5061–5079.

Walker, R., Moran, E., and Anselin, L. 2000. Deforestation and Cattle Ranching in the Brazilian Amazon: External Capital and Household Processes. World Development, 28(4):683–699.

Warren, R., de la Nava Santos, S., Arnell, N., Bane, M., Barker, T., Barton, C., Ford, R., Füssel, H.-M., Hankin, R. K., Klein, R., Linstead, C., Kohler, J., Mitchell, T., Osborn, T., Pan, H., Raper, S., Riley, G., Schellnhüber, H., Winne, S., and Anderson, D. 2008. Development and illustrative outputs of the Community Integrated Assessment System (CIAS), a multi-institutional modular integrated assessment approach for modelling climate change. Environmental Modelling & Software, 23(5):592 – 610.

Watts, R. D., Compton, R. W., McCammon, J. H., and Ouren, D. S. 2003. Intensity of Human Use, Backcountry Roads, and Analysis of Human Accessibility. In Group, T. R., editor, *Tahoe Research Group/Tahoe Environmental Research Center Publications*. Tahoe Research Group/Tahoe Environmental Research Center Publications, John Muir Institute of the Environment, UC Davis.

White, R. and Engelen, G. 2000. High-resolution integrated modelling of the spatial dynamics of urban and regional systems. Computers, Environment and Urban Systems, 24(5):383–400.

Whiteside, T. G., Boggs, G. S., and Maier, S. W. 2011. Comparing pixel-based and object-based classifications for mapping savannas. International Journal of Applied Earth Observation and Geoinformation, 13:884–893.

Xiang, W.-N. and Clarke, K. C. 2003. The use of scenarios in land-use planning. Environment and Planning B: Planning and Design, 30(6):885–909.

Yuan, Y., Smith, R. M., and Limp, W. F. 1997. Remodeling census population with spatial information from Landsat TM imagery. Computers, Environment and Urban Systems, 21:245–258.

Yue, T. X., Wang, Y. A., Liu, J. Y., Chen, S. P., Qiu, D. S., Deng, X. Z., Liu, M. L., Tian, Y. Z., and Su, B. P. 2005. Surface modelling of human population distribution in China. Ecological Modelling, 181 (4):72–88.

Zhang, Q., Justice, C. O., and Desanker, P. V. 2002. Impacts of simulated shifting cultivation on deforestation and the carbon stocks of the forests of central Africa. Agriculture, Ecosystems & Environment, 90(2):203–209.

Zhang, Q., Justice, C. O., Jiang, M., Brunner, J., and Wilkie, D. S. 2006. A gis-based assessment on the vulnerability and future extent of the tropical forests of the congo basin. Environmental Monitoring and Assessment, 114:107–121.

Zida, D., Sawadogo, L., Tigabu, M., Tiveau, D., and Odén, P. C. 2007. Dynamics of sapling population in savanna woodlands of Burkina Faso subjected to grazing, early fire and selective tree cutting for a decade. Forest Ecology and Management, 243(1):102 – 115.

Appendix A

Consolidation rules for PB-derived and OB products

Reconciliation rules between PB and OB approaches. Significant *p-values* at the 95 % confidence interval are in bold.

Table A.1: Consolidation of rules for PB-derived and OB products

Disagreement (PB-derived — OB)	*p-value*	Number of analyzed images	Best product	Final label
1. TC — HM	1.72 E-38	174	OB	HM
2. TC — LM	**1.10 E-91**	170	OB	LM
3. TC — NTC	**1.90 E-11**	158	OB	NTC
4. TC — WT	**3.57 E-3**	101	OB	WT
5. HM — TC	4.28 E-1	76	–	HM-TC
6. HM — LM	**9.11 E-3**	108	PB-derived	HM
7. HM — NTC	**1.56 E-12**	118	OB	NTC
8. HM — WT	**3.03 E-3**	72	OB	WT
9. LM — TC	2.83 E-1	44	–	LM-TC
10. LM — HM	1.81 E-1	16	–	LM-HM
11. LM — NTC	3.73 E-1	53	–	LM-NTC
12. LM — WT	1.23 E-1	2	–	LM-WT
13. NTC — TC	**7.46E-6**	97	PB-derived	NTC
14. NTC — HM	9.82 E-2	67	–	NTC-HM
15. NTC — LM	**3.22 E-4**	95	OB	LM
16. NTC — WT	4.86 E-1	39	–	NTC-WT
17. WT — TC	3.24 E-1	78	–	WT-TC
18. WT — HM	4.93 E-1	23	–	WT-HM
19. WT — LM	**1.19 E-2**	22	PB	WT
20. WT — NTC	**2.27 E-3**	54	OB	NTC