Spatial Planning, Growth, and Flooding—Contrasting Urban Processes in Kigali and Kampala

Eduardo Pérez Molina

SPATIAL PLANNING, GROWTH, AND FLOODING—CONTRASTING URBAN PROCESSES IN KIGALI AND KAMPALA

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. T.T.M. Palstra,
on account of the decision of the Doctorate Board,
to be publicly defended
on Tuesday, October 30, 2019 at 14.45

by

Eduardo Pérez Molina
born on October 11, 1982
in San José, Costa Rica

This dissertation is approved by:

prof.dr.ir. M.F.A.M. van Maarseveen (supervisor)
prof.dr. R.V. Sliuzas (supervisor)
prof.dr. V.G. Jetten (co-supervisor)

ITC dissertation number 369
ITC, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN: 978-90-365-4883-0
DOI: 10.3990/1.9789036548830

Cover design by Benno Masselink (based on the painting Panoramic View of a River with Low-lying Meadows by Jan van Goyen, circa 1644, courtesy of the Rijksmuseum Rijksstudio)

UNIVERSITY OF TWENTE. | ITC

Graduation committee

Chair	
prof.dr.ir. A. Veldkamp	Universiteit Twente
Supervisors	
prof.dr.ir. M.F.A.M. van Maarseveen	Universiteit Twente
prof.dr. R.V. Sliuzas	Universiteit Twente
co-supervisor	
prof.dr. V.G. Jetten	Universiteit Twente
Members	
prof.dr.ir. A. Stein	Universiteit Twente
prof.dr. K. Pfeffer	Universiteit Twente
prof.dr. A. van Rompaey	Katholieke Universiteit Leuven
prof.dr. P.H. Verburg	Vrije Universiteit Amsterdam
dr. S. Lwasa	Makerere University

To Andrea and Ana Victoria

With full faith in modern engineering and little respect for the power of nature, society continued to build within the floodplain. When the river overflowed its newly engineered banks, even more development lay in its path.

Robert W. Adler, *Restoring the Colorado River Ecosystems*

Summary

Urban growth is a factor known to intensify local flooding. By orienting urban development, land use planning may contribute to reduce flood risk through regulatory constraints. Two case studies were developed to determine the extent to which such strategy may be effective: Kigali, Rwanda (where land use regulations are stringently applied) and Kampala, Uganda (with much less effective institutions but important infrastructure investments over the last decade). Both cities are mid-sized (one to two million inhabitants), they share a physical context of hilly terrain and low-lying flood prone valleys but with divergent policy and institutional organizations.

Two main hypotheses were investigated based on the case studies. The relations between the physical system, through recurrent flooding, and the human settlement pattern were first explored. Urban growth is one cause of increased flooding but, in turn, flooding was thought to contribute to the urban pattern's evolution. Secondly, and based on this premise, a land use management system (with regulation a prominent component) was proposed as a flood risk mitigation strategy: these questions hinged around the feasibility of land use controls in the specific context of the cases (mid-to-large cities in Sub-Saharan Africa) and of their cumulative impact over the long run.

Spatially explicit prospective simulations of urban growth, up to the year 2030, were developed for both Kampala and Kigali to understand the impacts of flooding and land use regulations; additionally, a set of scenarios for Kampala was specified to explore the potential feedback effect between exposure to recurrent flooding and urban development patterns. The main lessons derived from these simulations were: in Kampala, which has until the present expanded without strong land use controls, the implementation stringent land use regulations (envisioned already in their strategic plans) would likely result in a more compact growth; however, in Kigali, the land use plan may have the unintended consequence of promoting sprawling patterns. Kigali was revealed to be a smaller urban system than Kampala, with the transitory benefit of not being yet impacted by recurrent flooding due to the scale of processes configuring urban growth. As for Kampala, (1) while land use planning may reduce exposure to flooding, it is unlikely to impact runoff generation and (2) explicitly incorporating feedback between flooding

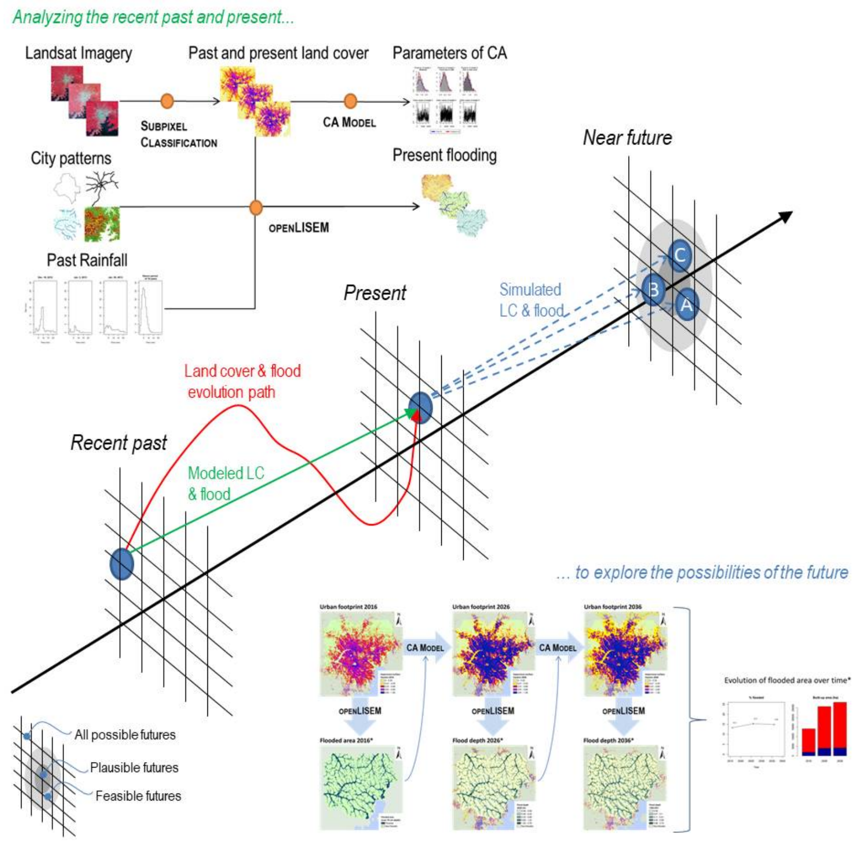
and urban growth makes visible a difference introduced by land use planning: under trend (unplanned) expansion, exposure to flooding is unlikely to constrain urban growth; however, under the double restriction of recurrent flooding and land use controls, much less new development is exposed to flooding.

The scenarios were carried out using a cellular automata of urban growth, specifically designed to be integrated with the flood model (implemented in *OpenLISEM*). Important characteristics of the model included: a continuous response variable, in the form of a land cover fraction value (built-up fraction for urban development but also vegetation, bare soil, and water fractions to complete the description of the landscape), a suitability based allocation procedure to mimic urban agents' locational preference, and the potential to explicitly account for several supply scenarios (which was especially important when considering the relation between population growth and densification in the scenarios). The suitability index was defined by a neighborhood effect, accounting for the immediate context of each potential development location, as well as ancillary variables representing accessibility, physical characteristics (slope and wetlands location), and informal settlements location.

The cellular automata model was developed using the Upper Lubigi sub-catchment of Kampala. The model was then expanded, calibrated, and validated for the metropolitan areas of Kampala and Kigali. Calibration was based on the application of the Metropolis-Hastings algorithm to determine the relative importance of each factor in the suitability index and using the land cover maps to simulate potential supply. Simulations using 2000 (for Kigali) and 2001 (for Kampala) as baseline years were generated for a 15 year period; for each simulated time step, landscape metrics were calculated. An intermediate year for which independent land cover maps were available (2009 for Kigali, 2010 for Kampala) was used for validation. The calibration approach proved useful in producing patterns that better replicate the evolution of urban growth patterns, relative to random parameters and data. However, some degree of equifinality was discovered in the model, since the uncertainty introduced by parameters was found to be less important than the amount of information (relevant spatial determinants) when validating the model.

The scenario assumptions on the presence of a feedback (for Kampala) and on the possible efficiency of land use controls were based on statistical analysis. The potential impact of flooding on urban growth was investigated using a structural equations model of Kampala, by making strong assumptions on the causal structure and performing confirmatory analysis to test whether data conform to such assumptions. The conclusion was, as expected, of a significant but weak restriction of flood impacts on urban growth patterns. Land cover maps were used to describe urban growth, the flood impacts were derived from the *OpenLISEM* model of Kampala. The effects of land use regulations on urban development patterns were calculated for Kigali using a difference-in-differences estimator. Kigali was chosen because already a stringent land use control system has been in place for the better part of two decades which, coupled with relatively rapid urban growth, provides a quasi-experimental setting. The conclusion was land use controls had in fact a statistically significant and strong impact on urban development.

In synthesis, this dissertation has developed a spatially explicit methodological framework to simulate the relations between urban growth, land use planning, and recurrent flooding. It was applied to two case studies in Sub-Saharan Africa, the cities of Kampala and Kigali. The framework is based on the integration of a cellular automata of urban growth and a flood model to reproduce the processes configuring spatial patterns. Scenarios, in turn, were specified based on the result of spatial statistical analysis of the relation between the main variables being explored. The results underscore the importance of opportunities but also the pitfalls of land use regulation as a policy response for flood mitigation.



Time line based on Allan Shearer (ed.) 2009. Land Use Scenarios. Environmental Consequences of Development

Samenvatting

De stedelijke groei is een factor die bekend is vanwege het feit dat het lokale overstromingen intensiever maakt. Door het oriënteren van stedelijke ontwikkeling, kan het plannen van landgebruik bijdragen aan de vermindering van het risiko van overstroming door gereguleerde beperkingen. Twee studies van gevallen werden ontwikkeld om te bepalen in hoeverre zo'n strategie effectief zou kunnen zijn: Kigali, Rwanda (waar de reguleringen van het gebruik van land strict worden toegepast) en Kampala, Uganda (met veel minder effectieve instellingen maar belangrijke investeringen in infrastructuur gedurende het laatste decennium). Beide steden zijn van middelmaat (één tot twee miljoen inwoners), zij delen een fysieke context van bergachtig terrein en laagliggende, aan overstromingen onderhavige valleien, maar met een afwijkende politiek en institutionele organisaties.

Twee van de belangrijkste hypothesen werden onderzocht op basis van studies van gevallen. De relaties tussen het physieke systeem, door herhaaldelijke overstroming, en het patroon van vestiging van mensen, werden eerst onderzocht. Stedelijke groei is één oorzaak van toenemende overstroming, maar aan de andere kant werd gedacht dat overstroming bijdroeg aan de ontwikkeling van de patronen van stedelijke groei. In de tweede plaats, en gebaseerd op dit uitgangspunt, werd een beheringssysteem van landgebruik voorgesteld (waarin regulatie een vooraanstaand component is) als een strategie tot vermindering van vloedrisiko: deze vragen hingen af van de haalbaarheid van de controles van landgebruik in de specifieke context van het geval (middelmaat tot grote steden in Sub-Sahara Africa) en van hun accumulatieve impact op de lange termijn.

Ruimtelijk expliciete prospectieve simulaties van stedelijke groei tot aan het jaar 2030 werden ontwikkeld zowel voor Kampala als Kigali om de impacten te begrijpen van overstroming en landgebruik regulaties; bovendien was een set scenarios gespecificeerd voor Kampala om het potentiële feedback effect te onderzoeken tussen blootstelling aan herhaaldelijk overstromen en stedelijke ontwikkelingspatronen. De belangrijkste lessen voortkomende uit deze simulaties waren: in Kampala, welke tot op heden is uitgebreid zonder sterke controles van landgebruik, de implementatie van stricte regulaties van landgebruik (al voorzien in hun strategische plannen) zou waarschijnlijk eindigen in meer compacte groei; in Kigali echter, zou het landgebruikplan het onbedoelde

gevolg hebben van de promotie van uitgestrekte patronen. Kigali bleek een kleinere stedelijke systeem te zijn dan Kampala, met het tijdelijke voordeel dat het nog geen impact had gehad van herhaaldelijke overstroming vanwege de schaal van processen die stedelijke groei configureren. Wat Kampala betreft, (1) terwijl het plannen van landgebruik de blootstelling aan overstroming verminderen, is het onwaarschijnlijk dat het de generatie van afvoerwater impacteert, en (2) de expliciete opname van feedback tussen overstromen en stedelijke groei maakt een verschil zichtbaar welke geintroduceerd wordt bij het plannen van landgebruik: bij trend (niet geplande) uitbreiding, is het niet waarschijnlijk dat blootstelling aan overstroming stedelijke groei tegenhoudt; echter, onder de dubbele beperking van herhaaldelijke overstroming en de controles van landgebruik, wordt veel minder nieuwe ontwikkeling blootgesteld aan overstroming.

De scenarios werden uitgevoerd met het gebruik van cellulaire automata model van stedelijke groei, specifiek ontworpen om geïntegreerd te worden met het overstromingsmodel (geïmplementeerd in *Open-LISEM*). Belangrijke kenmerken van het model waren: een continue responsvariabel, in de vorm van een gradatiewaarde voor landbedekking (opgebouwde fractie voor stedelijke ontwikkeling maar ook vegetatie, kale grond en waterfracties om de omschrijving van het lanschap de voltooien), een allocatieprocedure gebaseerd op geschiktheid om de locatievoorkeur van stedelijke agenten na te doen, en het potentieel om expliciet rekening te houden met verschillende aanbodscenario's (hetgeen specifiek belangrijk was bij het in acht nemen van de relatie tussen volksgroei en de verdichting in de scenarios). De geschiktheidindex werd bepaald door een buurteffect, goed voor de onmiddellijke context van iedere potentiële ontwikkelingslocatie, evenals aanvullende variabelen die de toegankelijkheid, fysieke kenmerken (locatie van hellingen en wetlands) vertegenwoordigen alsmede de locatie van informele vestigingen).

Het cellulaire automata model werd ontwikkeld met gebruik van de Upper Lubigi sub-stroomgebied van Kampala. Het model werd toen uitgebreid, gekalibreerd, en gevalideerd voor het grootstedelijk gebied van Kampala en Kigali. De kalibratie was gebaseerd op de toepassing van het Metropolis-Hastings algorithme om het relatieve belang te bepalen van iedere factor in de geschiktheidindex en door het gebruik van de land oppervlaktemappen om potentiële bevoorrading te simuleren. Simulaties die 2000 (voor Kigali) en 2001 (voor Kampala) als basisjaren gebruikten, werden gegenereerd voor een periode van 15 jaren; voor iedere gesimuleerde tijdstap werden landschapsstatistieken berekend. Een tussenjaar waarvoor onafhankelijke land oppervlaktemappen beschikbaar waren (2009 voor Kigali, 2010 voor Kampala) werd gebruikt voor validatie. De kalibreringsbenadering bleek nuttig te zijn bij het produceren van patronen die beter de evolutie dupliceren van stedelijke groeipatronen, relatief aan willekeurige parameters en data. Er werd echter een graad van equifinaliteit ontdekt in het model, aangezien de onzekerheid ingevoerd door parameters minder belangrijk bleek te zijn dan de hoeveelheid informatie (relevante ruimtelijke determinanten) bij het valideren van het model.

De scenario veronderstellingen voor wat betreft de tegenwooordigheid van een feedback (voor Kampala) en omtrent de mogelijke efficiëntie van controles van landgebruik, waren gebaseerd op statistieke analyse. De potentiële impact van overstroming op stedelijke groei werd onderzocht met behulp van een structureel vergelijkingsmodel van Kampala, door sterke veronderstellingen te maken omtrent de causale structuur en door bevestigingsanalyses te maken om te testen of de data overeenkomen met die veronderstellingen. Zoals verwacht, was de conclusie dat van een belangrijke maar zwakke beperking van overstromings impacten op stedelijke groei patronen. Land oppervlakte kaarten werden gebruikt om stedelijke groei te omschrijven, de overstromings-impacten werden afgeleid van het *OpenLISEM* model van Kampala. De effecten van de regulaties van landgebruik op stedelijke ontwikkelingspatronen werden berekend voor Kigali met het gebruik van een verschil-in-verschillen schatter. Kigali werd gekozen omdat er al een stricte landgebruik controle systeem bestond voor het grootste gedeelte van 2 decennia, welke, gekoppeld aan relatief snelle stedelijke groei, een quasi-experimentele achtergrond verschaft. De conclusie was dat controles van landgebruik in feite een statististisch balangrijke en een sterke impact hadden op stedelijke ontwikkeling.

In synthese heeft dit proefschrift een ruimteijk epliciet methodologisch kader ontwikkeld om de relaties te simuleren tussen stedelijke groei, landgebruiksplanificatie en herhaaldelijke overstroming. Het is toegepast op twee case-studies in Sub-Sahara Afrika, de steden Kampala en Kigali. Het kader is gebaseerd op de integratie van een cellulair automata model van stedelijke groei en een overstromingsmodel om de processen te reproduceren die ruimtelijke patronen configureren. Scenarios, aan de andere hand, werden gespecificeerd, gebaseerd op het resultaat van ruimtelijk statistische analyse van de relatie tussen de belangijkste variabelen die onderzocht werden De resultaten onderlijnen het belang van gelegenheden, maar ook de valkuilen van de landgebruikregulatie als een politiek antwoord voor de verzachting van overstroming.

Acknowledgments

Over the course of the past five years, as the process which ultimately lead to this report unfolded, I have had the good fortune of sharing with a group of people who have encouraged and improved my thinking on the research – and who made for me a happy life in Enschede.

First and most especially, I would like to acknowledge my supervisors, prof.dr. Martin van Maarseveen, prof.dr. Richard Slizuas, and prof.dr. Victor Jetten. Not only have they been a source of unwavering support, insight, and expertise (on the science and on the specific context of Sub-Saharan Africa), they have also contributed over these years to educate me into a better researcher. Certainly this dissertation would not have been possible without their patience, and sage advice.

I must also express my gratitude to Yakob Umer, colleague from the Department of Earth System Analysis, for developing and facilitating the flood model of Kampala I have used throughout this dissertation.

At the risk of my imperfect memory misguiding me into an incomplete list, I do feel the need to express my appreciation to my colleagues of the PGM Department at ITC: prof.dr. Yolla Georgiadou, for her generous advice over the years and for organizing a space for reflection on geo-technology and policy that contributed to sharpen my thinking on the suject; prof.dr. Karin Pfeffer and dr. Mafalda Madureria, organizers of the PhD seminar where, similarly, some enlightening ideas came to be discussed; prof.dr. Anne van der Venn, who since chairing my MSc proposal up till his retirement was a source of far-sighted counsel (indeed, I think one of my main regrets, looking back, was my inability to act on these earlier – alas, such is the PhD process). I must also thank my module 8 and 12 colleagues, Mark, Frans, Anna, Monika (with who I thought through the idea of landscape metrics as related to urban growth modeling), and Nina; Cheryl and Rogier, teammates (also with Mark) in developing the project of the Spatial Engineering module; and Diana, colleague on the CARE Project, as well as (once again) Richard for involving me in all of these; Andre for so many stimulating (work and non) conversations; Jeroen who led so many interesting discussions on the MSc course; and of course the essential Petra. As many ideas were freely discussed and offered, I can't conceive of this project without all of you.

The fieldwork stage was possible due to the support of numerous colleagues in Rwanda and Uganda, many ITC alumni, who were generous with their time and resources. In Kigali, I gratefully acknowledge the help of Fred Mugisha, Hervè Habonimana, Marc Manyfika, Ignace Sendama, Emmanuel Ingabire, and Alphonse Nkurunziza; and in Kampala, I benefited from interviewing and information provided by Zabali Flavia Musisi (GIS section of Physical Planning) and Justus Akankwasa (Engineering), both from the KCCA.

On the home front, it was Gustavo and Tatjana who kept me fed, warm, and mentally stable for the better part of three years: I couldn't have asked for better housemates. So many years, so many friends whose generosity made this possible: my fellow PhDs Manuel, Alby and Sheila, Abhi, Sara, Razzieh, Shima, Andrés, Valentina, Alice, Rosa, Lia, Marleny, Zahra, Adriana, Effie, Andrea and Fernando, Caroline, Claudia, Kwabena, Maria Fernanda and Alain, Irene, Fenna, Sergio, Deepshikha, Sven; Dyviani, one of my guardian angels, Parya, Ana and Inês, Emma, John and Lorena, Verónica, Jiong, Nadia, Alicia, Gerson, Lis and Inti, Lucas, Siddhi, Alfredo V., Ana Patricia, Diana, Marisol, Lilliana, Laura, Lina, Alfredo C., Rohan, Dimo... The list is long and if I missed anyone, please accept the apology of a brain that is recovering (for, while mostly good, this research was *tiring*!)

I must also gratefully thank colleagues, friends, and mentors from back home, who provided the impulse to start, follow, and finish this research project: prof.dr. Rosendo Pujol and Jonathan Agüero (who brought me back to Costa Rica), from the University of Costa Rica, and dr. Juan Carlos Vargas of Boston. The University of Costa Rica funded four months for the development of this project.

And lastly, because they are first in my heart, my two A's: Andrea and Ana V., you kept me going when otherwise I would have stopped. Were it not for you, this would have never been completed.

Contents

List of Figures

List of Tables

1 Setting the scene

1.1 A rationale for looking at the urban environment of Sub-Saharan Africa

World population estimates for 2015 attested to the global importance of Sub-Saharan Africa: one in seven human beings lived in the subcontinent, nearly one billion people, around 40% of them in cities (United Nations, 2018). And yet, despite a decade and a half of solid economic growth, the number of poor increased in Africa over that period (Signé, 2018). Poverty being a very complex social phenomenon, a myriad of factors have contributed to this state of affairs. Signé (2018), though, emphasized three that are especially challenging in the context of Sub-Saharan Africa1: pro-poor policies have often been ineffective, demographic growth has been too rapid, and economic growth did not create quality jobs.

Cities have been proposed as a solution to the economic productivity quandary and its associated jobs problem long since before the advent of regional economic theory, although arguably, in Sub-Saharan Africa, they have not been successful in triggering economic development (Lall, 2017). As argued by Baptista (2003) by means of macroeconomic modeling, greater densities of economic activity, such as those characterizing cities, have a multiplicative positive effect on economic output caused by the agglomeration of economic agents. Densification in the broad sense (of economic activities, of workers, of population), however, also poses its own challenges. Baptista (2003) includes in his model a congestion elasticity as well as an agglomeration elasticity to account for these. The fundamental tenet of regional economics, then, can be boiled down to a trade-off between the negative externalities of congestion and the positive externalities of agglomeration.

Sub-Saharan Africa has been undergoing rapid urbanization with manufacture and service sectors development lagging (Gollin et al., 2016; Lall, 2017), to a certain extent mirroring trends already experienced in Latin America and east Asia (Glaeser and Xiong, 2017). While the experience from these other contexts suggests development in the broad sense will eventually follow urbanization, there is no denying the present reality of urban systems in Sub-Saharan Africa: that of the negative externalities of congestion outweighing the economic advantages of agglomeration effects. Dealing with these negative externalities requires effective and efficient provision of infrastructure (a result only possible with strong public institutions) and individual-level incentives to prevent overuse (Glaeser and Xiong, 2017). One may even argue previous economic growth was possible only when such infrastructure was supplied, as in the largest Latin American and east Asian cities.

There is, however, a more distinct difference between the urbanization of the past and that of the Sub-Saharan present. The cities of the industrialized north grew in a world of seemingly boundless resources; and even though the first discussions on the limits to growth (Meadows et al., 1972) took place simultaneously with the expansion of urban population in Latin America and east Asia, the paradigm shift towards sustainability came after most of the urbanization dynamic had played out. Sub-Saharan Africa, to the contrary, must seek the benefits of urbanization in a world constrained by the excessive consumption of humanity. This entails the need to consider the natural environment at the forefront of the human settlements creation process.

Within a large conceptual realm of environmental issues confronting urban systems in Sub-Saharan Africa, Douglas et al. (2008) identified flooding as a problem disproportionately affecting the urban poor and increasing in magnitude across Sub-Saharan Africa. They identified two main drivers: the physical growth of cities, which causes more impervious areas and hence greater runoff, and climate change making extreme events more frequent and more intense. Dodman et al. (2017) further argued the socioeconomic and institutional setting of urban Sub-Saharan Africa contributed to shape vulnerability, notably through urban expansion into exposed areas and lack of infrastructure; and indeed how the specific demographic (young societies), economic (informal and generally poor), and governance characteristics all constituted challenges to but also opened opportunities for risk mitigation.

Given this state of affairs, how can the generation of knowledge contribute to improve the livelihoods of residents in Sub-Saharan Africa's cities? The main problem addressed in this dissertation is to understand how land use planning responses can contribute to improve cities in Sub-Saharan Africa and the livelihoods of their residents through flood risk mitigation. Deeply embedded within this question is the methodological topic of how to generate basic information to design and evaluate policy responses in a data scarce environment.

1.2 Coupled human and natural systems in hydrology and regional science

The notion that human agency modifies the environment has long been acknowledged as a fundamental premise of land change science (Lambin et al., 2001). At regional scales, the analysis of these relations permits the integration of environmental transformations with the drivers that have directly caused them (Meyer and Turner, 1990): these processes are characterized by complex interactions between the human and the environmental elements, the environmental consequences of human actions variously feeding back into the actions themselves.

Coupled human and natural systems (CHANS) were developed as a framework to address precisely such problems of environmental transformation. As described by Liu et al. (2007), these systems present multiple reciprocal interactions between people and nature through complex feedback loops. They are often influenced by larger-scale external factors and their internal interactions, mediated by structures that arise within the system (for example, urban form and infrastructure in cities). These interactions may be non-linear. Thresholds – points at which a system changes into an alternate state – are a particularly relevant form of non-linearity: CHANS transform over time and over space. The analysis of such systems, therefore, crucially requires sufficient spatial and temporal extent and detail to elucidate the most relevant dynamics of the system.

CHANS is a specific instance of a framework to study what can be more generally termed socio-environmental research (Binder et al., 2013; Pulver et al., 2018). A first comparison of such frameworks, developed by Binder et al. (2013), suggested three main characteristics to distinguish among them: whether the relation between the social and environmental components went one or both ways, whether it was centered on the social or environmental perspective of the system, and whether the emphasis was on action or analysis. However, the rapid evolution of the field necessitated an update on these criteria (for example, all modern frameworks make use of bi-directional relations). Pulver et al. (2018) undertook further analysis, based on six frameworks (in addition to CHANS, they discussed human ecosystem framework, resilience, integrated assessment of ecosystem services, vulnerability framework, and social-ecological systems framework). They found two core features on all of these frameworks: first, the distinction between the social and the environmental elements, generally with an equal level of importance; and second, the central role of components (the building blocks), connections between components, scale, and context.

From this viewpoint, trade-offs exist between emphasis placed by different frameworks on the importance of these four elements. CHANS, in particular, is centered on connections rather than components: it loosely organizes the components into two main blocks (environmental and social) but emphasizes hierarchical connections (Pulver et al., 2018), facilitating the exploration of relations between different scales and across the socio-environmental divide. This fits comfortably with the urban flooding phenomenon, as the two main systems (the urban system and the flooding process) can be analyzed through decomposition and characterization of internal relations before exploring feedbacks and non-linearities between the human-environment divide. Additionally, unlike the human ecosystem framework (which emphasizes ecosystem management) or social-ecological systems (related mainly to common pool resources), CHANS is a systems approach, and thus not linked to any particular environmental (or social) phenomenon (Pulver et al., 2018).

Likely because of cellular automata's capabilities in replicating the spatially explicit fractal dimensions exhibited by urban dynamics (Batty, 2009) – and especially urban growth –, the bottom-up, emergent behavior of a system approach has long been in use by urban and regional planners. However, focus has been on the central mechanism of the Alonso-Mills-Muth model, the trade-off between accessibility and space (Glaeser, 2008). The prototypical structure of cities has been built around land use and transportation (concretely traffic, which represented the linkages within the system); see Batty (2009): the human environment (land use and buildings) interacting with human agents. This regional science domain of application has concentrated only on the human element, generally avoiding environmental dimensions (with the possible exceptions of fuel consumption and air pollution). It is, even if an important previous example, not a proper CHANS approach.

The application of CHANS in hydrological research is much more recent; as late as 2012, socio-hydrology was being proposed to operationalize CHANS and incorporate social elements into hydrological modeling (Sivapalan et al., 2012). Socio-hydrology as a field intends to deal with emergent phenomena from a diversity of coupled human-water systems, generally conceived as two-way feedbacks between human and natural elements and which are hypothesized to have common theoretical explanations (Pande and Sivapalan, 2017). Socio-hydrology does not seem to consider any emergent phenomena, even if the framework could be applicable to bottom-up dynamics; rather, it concentrates on problems that commonly manifest themselves across different contexts (such as levee infrastructure, flooding, and settlement patterns). It claims to be distinguished by the treatment of human agency as endogenous to the coupled system (Pande and Sivapalan, 2017).

Examples of socio-hydrological applications to flood phenomena include Chen et al.'s (2016), which describes theoretically – based on linked differential equations – how the environmental sensitivity of a society evolved from favoring flood (and channelization of wetlands) to conservation, as the wetlands were degraded. Their model predicts this sensitivity will swing back to neutral as wetlands recover, although increased rainfall could accelerate changes in social norms again favoring flood protection over ecological integrity. Di Baldassarre et al. (2013) created a more theoretically general model, also based on differential equations, linking settlement patterns and location decisions, societal awareness of flood risk, infrastructure and investments (specifically, levees to reduce flood occurrence), social psychological shock in response to a flood event, and the flooding itself; they were able to replicate through scenarios of high/low cost (of investment in levees) observed patterns of settlements choosing to "live with floods" and to "fight floods" (build levees), as well as the shift from frequent low magnitude events to larger, more damaging but less frequent floods, in response to levee development.

While socio-hydrology has advanced the introduction of social elements into the analysis, one must caution the program is very far from complete. A recent survey of applications (Xu et al., 2018) found a large majority of cases have been developed by hydrologists. This is manifested in a dominance of systems engineering and differential equations approaches, and questions posed in such a way as to deter social scientists from collaborating in the field.

Given its aspiration to quantitatively integrate models and the prospect of new evidence from spatially explicit analysis (Pande and Sivapalan, 2017), the combination of urban growth models and flood models may be a valuable direction in CHANS hydrological research. While conceptually straightforward, the integration of urban growth and flood models has been rare. Most case studies have analyzed the impact of prospective land patterns on hydrological or hydraulic outcomes, among which the work of Khan et al. (2018), Nigussie and Altunkaynak (2016), Ciavola et al. (2014), Huong and Pathirana (2013), Kumar et al. (2013), and Poelmans et al. (2011) are typical recent examples: the setup is for hydrological outcomes to be treated as a dependent variable, explained by land cover dynamics.

Khan et al. (2018) projected possible future trend conditions of Dhaka, Bangladesh, from the baseline of 2010 to 2050. The urban growth model of Dhaka was built using the Dinamica EGO model (Soares-Filho et al., 2002) and applying a memetic algorithm (Veerbeek et al., 2015) to calibrate the weight of different determinants (both historical land use and other maps). The outputs of this urban growth model were used jointly with the characteristics of a major flood event, which occurred in 2004 (and was simulated in a spatially explicit manner), to understand what damage would happen should a similar event take place again: they predicted a substantial increase in overall damage, as well as identifying key factors (particularly urban growth patterns, current and prospective) and uncertainties (such as the effect of climate change, deliberately excluded, or possible flood protection infrastructure, as well as possible changes to urban growth patterns).

Nigussie and Altunkaynak (2016) applied the SLEUTH model (Clarke et al., 1997) to Istanbul and generated four scenarios with increasing levels of exclusion (constraints on urban growth); the two most restrictive modeled the urbanization of a former military base, which impacts the Ayamama watershed within Istanbul. To assess the hydrological impact, they used HEC-1 to calculate the output hydropgraphs only for the Ayamama watershed – and found the projected urbanization of the military base, despite greater constraints, increases peak discharge and reduces time to peak for Ayamama. Ciavola et al. (2014) modeled economic and demographic trends to estimate future land demand and allocated it, also with the SLEUTH model; based on the scenarios, they estimated nutrient load and runoff for selected catchments of the Delaware-Maryland-Virginia, U.S. Kumar et al. (2013) developed a simple cellular automata model of Roorkee, India to simulate urban growth; they assessed the runoff impacts using the NSCS method. Huong and Pathirana (2013) simulated future land cover trends with the Dinamica EGO model (Soares-Filho et al., 2002), jointly with an atmospheric forcing model. Land cover simulations were used solely to explore the sensitivity of atmospheric conditions to land systems. Hydraulic flood routing was estimated assuming their study area as uniformly urban.

Poelmans et al. (2011) used a cellular automata-based model of urban growth to project the expansion of Brussels within the Moleenbeek catchment; they further specified scenarios of future climate (for 2050, based on the application of regional circulation models) and coupled these with rainfall-runoff and a 1D hydrodynamic model. Ultimately, they produced maps of flooded area and estimates of peak flow incorporating both climate change and urban growth as drivers.

Relatively few urban growth studies have explored the obverse of this modeling disposition, i.e. treat land dynamics as (partially) a consequence of potential (or actual) flood risk. In their study of La Paz, Mexico, Steinitz et al. (2005) used a normative buffer to frequently flooded but usually dry watercourses as a proxy of recurrently flooded areas and restricted their development for constrained development scenarios. Sakieh et al. (2017) designed a suitability-based urban growth model (with neighborhood land use modifying the suitability value) to explore polycentric urbanization of Gorgan, Iran; they compared an urban suitability layer with an environmental risk layer as guides for prospective allocation.

Interpreted within a CHANS framework, neither of these approaches fully satisfies the two-way feedback envisioned by the socio-hydrology perspective. The state of knowledge, with technical tools capable of but not used to explore a spatially explicit two-way feedback between human and natural subsystems, invites the reflection: can the integration of urban growth models and flood models be used to endogenize the boundary conditions, i.e. dynamically link the evolution of urban and flood patterns? Given the presence and influence of large uncertainties, inherent to spatial modeling, what can and cannot be learned from such systems and with these methodologies?

1.3 Research aims: urban flooding as a coupled human and natural system

The conceptualization of urban flooding analysis within a CHANS framework becomes relatively straightforward, as can be seen in figure 1.1, when formulated abstractly. As proposed in this dissertation, the core of the system is composed of three elements: the built-up land, the natural landscape (which includes natural ecosystems and elements, notably wetlands and undeveloped floodplains), and urban growth. The first two components describe the basic characteristics of a city and the latter, the main dynamic change to the spatial structure they jointly determine.

The physical reality of urban flooding is, as well, conceptually simple. The hydrological analysis of rainfall-runoff results in estimates of how much rainfall is infiltrated into the soil or intercepted and how much

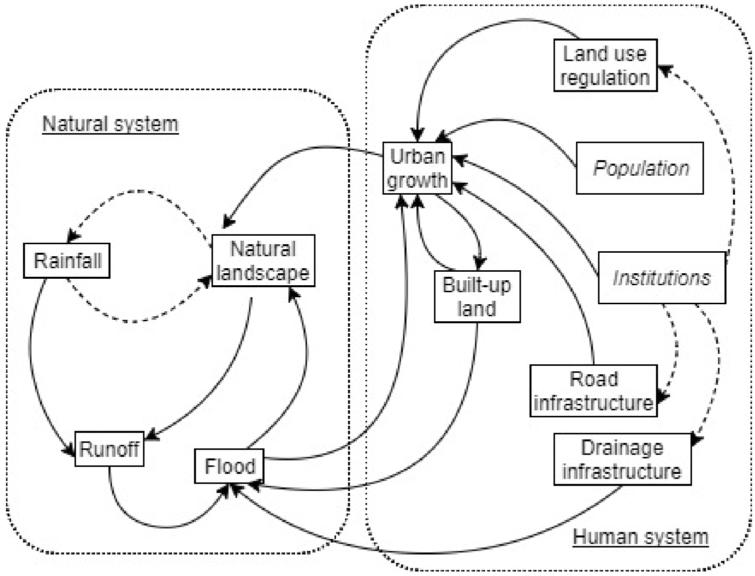


Figure 1.1 Conceptual framework: urban floods as CHANS

becomes water flowing over the landscape. Soil properties, such as infiltration, and vegetation characteristics are parts of the natural landscape. Sene (2010, p. 110) describes the physical processes: rainfall is either infiltrated into the soil, intercepted (mainly by vegetation), or flows over the landscape; terrain, through elevation and other characteristics, routes the water flowing over the surface. Changes in land cover patterns are known to affect the flood patterns: as more areas are urbanized, imperviousness increases (and so does the fraction of rainfall flowing over the landscape), resulting in more and faster flooding (Smith and Ward, 1998). Translated into the terms of the system proposed in figure 1.1: the amount of rainfall and the physical characteristics of the natural landscape determine how much water becomes infiltrated and how much runs off the terrain, a portion eventually becoming flood. By sealing the surface of the soil, the buildings that are part of the built-up land increase the amount of runoff.

The social dynamic is perhaps more complex, as it cannot be fully understood without reference to the natural environment nor with so few elements: (1) urban growth increases the amount of built-up land, and decreases that of the natural environment; (2) the amount and location of urban growth are determined by population growth and by societal preferences, some physical (including the characteristics of the built-up land, feeding back into urban growth) and others normative: the values and institutional actors through which society decides on regulations and infrastructure, as well as taxation and attitudes toward natural systems; in particular, (3) land use regulation re-distributes the location of urban growth and may constrain its amount, (4) some infrastructure, specifically drainage infrastructure, reduces flooding by deliberately modifying the environment, (5) road infrastructure attracts urban development.

Note of figure 1.1 that the connections of interest within the modeling of this dissertation are shown as solid lines, other long run connections – notably the relations between land and climate and between society's institutions (in the broadest sense of including institutional actors but also the societal norms and values that empower them) and the main instruments of the land use planning system (infrastructure investment and land regulation) – not directly modeled are represented as dotted lines.

A CHANS framework, applied to urban flooding, collates the need of accessible land for urban development and the flood impacts of surface sealing in the landscape. From these conditions follow the two research questions around which this dissertation is organized, namely: (1) *How do urban growth and flood processes interact?* (2) *To what extent can land use planning contribute to mitigate the impact of floods on urban spatial patterns?* Re-stated in terms of the conceptual framework (figure 1.1), special interest is placed by this structuring of the research questions on the cross-connections between the natural system and the human system, both direct (the relations between urban growth and flood) and indirect (the effects of infrastructure, the product of a societal process, on both flood and urban growth).

Within this scope, the following objectives were defined to operationalize the study:

To quantify the relationship between land cover change and flood hazard for selected case studies.

1. To generate land cover data models appropriate for reflecting urban dynamics and as inputs for flood modeling.
2. To analyze land cover change and flood patterns and their main determinants.
3. To design and calibrate a modeling tool that can replicate existing land cover and flood patterns, and simulate alternative conditions in metropolitan contexts.

To explore the potential of land use regulation and infrastructure investment in reducing the impact of pluvial flooding.

1. To formulate scenarios for selected metropolitan areas incorporating land use regulation.
2. To simulate, using the developed tool, the defined scenarios and assess the impacts of flooding on the urban system.

These questions are tackled through the deployment of spatially explicit quantitative models that describe the non-linear relations relating most elements in figure 1.1: the development of a cellular automata model of urban growth (describing the evolution of the human and natural environment as urban growth occurs) and the applicationof the openLISEM rainfall-runoff model (Jetten, 2018), which converts the characteristics of the environment and of a rainfall event into flooding. Societal effects – population growth and the norms and values behind regulation and infrastructure aspects of land use planning – are explored by means of scenario planning and the comparison between two case studies (as one selected case study, Kampala, is prototypical of a Sub-Saharan land use planning system, with weak institutions and ineffectual implementation of plans, in contrast to the second case study, Kigali, an example of a very stringent application of land use regulations; see subsection 1.4.2).

1.4 Methodological approach

1.4.1 Building a CHANS model for urban flooding

The practice of model development as a means of understanding and of acting upon the world must be purposeful and deliberate (see van Vliet et al., 2016, Magliocca et al., 2015, and the references therein, specifically the sections on the model development cycle). Yet a clear definition of intent is but the first step in an often meandering process that should lead to the understanding synthesized in the model itself.

Rosenblueth and Wiener (1945) described the construction of knowledge as starting with a *closed box* relating *inputs* to *outputs*. Several alternative configurations could, conceivably, relate the same inputs/outputs. Scientific progress is achieved by gradually opening the closed boxes and determining their inner configuration. This view coincides with contemporary notions on the design of models (concretely, agent based models, although they are valid in general) embodied in the KISS approach (the "keep it simple stupid" of military fame). KISS-based models begin relating the minimum amount of variables through the least possible amount of mechanisms (Edmonds and Moss, 2005). New ones, both variables and mechanisms, are subsequently added *as needed* (Edmonds and Moss, 2005) – which is to say, closed boxes are progressively opened.

How complex should the model ultimately be? (How many boxes should be opened, how many variables and mechanisms added?) The principle of *parsimony* (Seasholtz and Kowalski, 1993) posits as few as possible to reach, from given inputs, a certain level of accuracy in the prediction of outputs (Pitt et al., 2002). Given input and output data for a single case, initially the addition of complexity to the model increases both generalizability – the capacity of the model to predict outputs for a new, unseen sample of inputs – and goodness-of-fit (how good the model is in predicting the originally given output from the given input). As model complexity increases further, beyond a certain level so does goodness-of-fit but not generalizability (this is because the model begins to fit *noise*, i.e. random error, in addition to *signal* in the data, a phenomenon known as overfitting); indeed, generalizability eventually decreases as uncertainty becomes more important than the meaning in the data (Pitt et al., 2002). Thus, all else held equal, one should prefer simpler models to more complex ones.

Yet all things are never equal: a complex model opens more boxes – offers more knowledge – than a simple one, provided it is generalizable. Hirschman (1985) elegantly developed this argument in an essay aptly titled “Against Parsimony: Three Easy Ways of Complicating some Categories of Economic Discourse”: his argument, economic theory had simplified human behavior to the point that phenomena amenable to theoretical analysis had been excluded from traditional economic science. He thus illustrated the dangers of oversimplification against which Einstein, in a famous maxim, warned, “A scientific theory should be as simple as possible, but no simpler” (cited in Jørgensen and Bendoricchio, 2001, page 51). One should also note the arguments of Edmonds and Moss (2005), cited here to describe the KISS paradigm but who, in fact, summarize this approach as a starting point to criticize simplicity for simplicity’s sake in agent based model development.

Sources of increasing model complexity in the context of this dissertation are: feedback effects, static vs. dynamic relations, and the number of key variables (see table 1.1 for an overview).

1.4.2 Study Areas: Kampala, Uganda and Kigali, Rwanda

Cities of Sub-Saharan Africa offer a valuable opportunity to study the relations between urban growth, flooding, and spatial planning. Sub-Saharan Africa is undergoing a process of rapid urbanization with annual population growth rates of larger cities over 3.00% (Cobbinah et al., 2015). This accelerated growth has resulted in a substantial increase of of urban footprints in many cities of Sub-Saharan Africa over the last 15 years (Sliuzas, 2004; Vermeiren et al., 2012; Nduwayezu et al., 2016; Hou et al., 2016; Pesaresi et al., 2016). Many cities in Sub-Saharan Africa present a tropical climate characterized by highly localized, intense, and short rainfall events that typically lead to rapid flooding; generally, climate patterns show high variability between and within years (Douglas et al., 2008).

The cities of Kampala, Uganda and Kigali, Rwanda were selected as case studies for this dissertation. As noted by Goodfellow (2013a), both cities share many characteristics – overall population (between one and three million residents) and rates of urbanization, physical environment (hilly terrain with large wetland areas), political regimes, and structure of the national economies. However, they have diverged in one crucial aspect: Kigali has opted, since circa 2009, for a land use planning system based on the stringent application of regulation; Kampala has tolerated a much more chaotic situation (Goodfellow, 2013a) but with a recent history of systematic investment in large urban infrastructure projects (roads and drainage).

The choice of study areas, therefore, provides sufficient variation in terms of urban development, climate, and policy to fruitfully explore the urban flooding as a CHANS, as hypothesized.

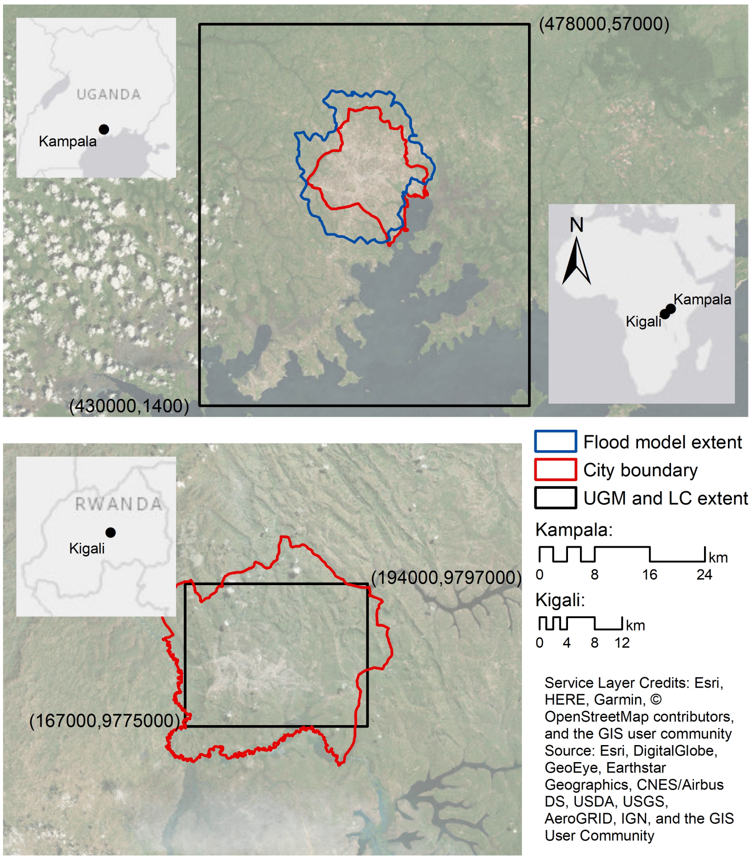


Figure 1.2 Study areas: metropolitan areas of Kampala, Uganda and Kigali, Rwanda

Kampala, Uganda

Kampala (see figure 1.2) is Uganda's largest city and its major center of commerce, as well as being – with two a half million residents – one of the largest city in East Sub-Saharan Africa (United Nations, 2018). The Kampala metropolitan area, located at latitude $0°19'N$ and longitude $32°35'E$, covers approximately $325km^2$. The city proper is divided into five districts, although a significant part of new development occurs beyond the city boundaries. Kampala is located north of Lake Victoria, in a hilly terrain that also contains large areas of wetlands. Its tropical weather and soil infiltration properties already lead to large runoff volumes, a main cause of recurrent flooding, which has been exacerbated by urban growth (Sliuzas et al., 2013).

Kampala has expanded, over the last twenty years, at a very fast rate. Abebe (2013) reported its urban footprint expanded from $117.0km^2$ in 1995 to $324.6km^2$ in 2010, a threefold increase. Vermeiren et al. (2012) estimated the city had experienced exponential growth since the 1970s. These figures mirror the recorded population increase, from 964 thousand in 1995 to 2.6 million in 2015 (United Nations, 2018): an annual growth rate of over 5.0%.

Kigali, Rwanda

Kigali, Rwanda (see figure 1.2) is a city of around one million residents (United Nations, 2018), the largest urban agglomeration and administrative center of Rwanda. Kigali is located near the Equator, at coordinates $1°57'S$ $30°4'E$, in the Central Plateau, a large region of hills and valleys formed by the folding and erosion of soil strata (Habonimana et al., 2015). The Nyabugogo river is located to the north-west of Kigali. While recurrent flooding is less frequent than in Kampala, perhaps because the city is smaller and has until recently developed at a slower pace, important areas within Kigali show a high propensity to flooding and this hazard was incorporated into land use regulations to mitigate its impacts (Bizimana and Schilling, 2010).

Kigali suffered considerable dislocation as a consequence of the 1994 genocide and its aftermath. Since then, however, it has exhibited the fast paced growth typical of Eastern Sub-Saharan Africa (in part due to returning refugees; see Goodfellow & Smith, 2013). Nduwayezu et al. (2016) reported Kigali's urban footprint increased from $42.1km^2$ in 1999 to $95.5km^2$ in 2014, more than doubling in 15 years. The total population of Kigali, during 2000-2015, rose from 498 thousand to 951 thousand (United Nations, 2018), an annual growth rate of 4.4%.

1.4.3 Dissertation outline

This dissertation can be thought of as the development of an increasingly complex coupled urban growth and flood model, ultimately used to simulate integrated scenarios of rainfall, policy, and demographic conditions. Chapter 2 reports on the development of land cover data models, which are a static representation of urban growth in various moments in time. Chapter 3 makes use of statistical techniques to unpack the feedback relation between flood and urban growth. Chapters 4 and 5 develop an urban growth model based on cellular automata. Chapter 4 summarizes the development of a prototype model for the Upper Lubigi catchment of Kampala, chapter 5 improves and extends this model and develops techniques to calibrate it for Kampala and for Kigali. Chapter 6 examines the effects of land use planning on urban patterns in Kigali. Chapter 7 unifies all methods by integrating the flood model with the urban growth model and by using them to simulate prospective scenarios that compare trend conditions to alternative future patterns envisioned in the land use plans of each city.

Table 1.1 shows how the different models reported are organized in terms of complexity. Generally, the links between land cover and flood were first established and then land use planning elements were introduced. Furthermore, the results from simpler static models informed

Table 1.1 Organization of models to explore the relations between spatial planning, urban growth, and flooding

	Static	Dynamic
$A \Rightarrow B$	**Ch. 2**: B = land cover A = reflectance (physical variable) Technique: Spectral mixture analysis and boxplots **Ch. 6**: B = land cover A = land planning instruments Technique: difference-in-differences model	**Ch. 4 & 5**: B = land cover A = spatial determinants Technique: cellular automata and Monte Carlo for calibration
$A \Leftrightarrow B$	**Ch. 3**: B = land cover A = flooding Technique: Structural Equations Modeling	**Ch. 7**: B = land cover A = flooding Technique: cellular automata and flood model Simulation of rainfall Scenarios of development supply (including land use plans)

more complex dynamic ones and the results from models assuming exogeneity influenced feedback models. An advantage which follows from access to two case studies is that, at different stages of the model development, the most appropriate case can be used to generate a specific component. However, there are also problems of interpretation when comparing both cases: firstly, since methodologies often need case-specific adjustments, should one attribute the differences between cases to a substantive phenomenon or to the methods?; secondly, and closely linked to the former: when a result is derived for one case study, to what extent can one extrapolate it to the other case study? Chapter 4 was a follow up to Sliuzas et al. (2013) and, because of this, limited to the Upper Lubigi subcatchment. Similarly, chapter 3 only analyzed Kampala, because it is there where the relation between recurrent flooding and urban growth is more apparent. Chapter 6 only concerns Kigali because it is there where one can find a strict application, and therefore a potential effect, of land use planning instruments. All other chapters (2, 5, and 7) were developed for the metropolitan region of both Kampala and Kigali.

2 Sub-pixel land cover maps to explore urban dynamics in Sub-Saharan Africa

Abstract

The methodology to and results of generating sub-pixel land cover classifications for two time series (Kampala, 2001-2016 and Kigali, 2000-2015) of impervious surface, vegetation, soil, and water are presented and discussed. The estimation of sub-pixel land cover fractions was achieved through the application of supervised and unsupervised classification methods, the results of which are matched to a conceptual model of land cover.

Keywords: Urban expansion, spectral mixture analysis, Informality, Flooding, Kigali (Rwanda), Kampala (Uganda)

2.1 Introduction

Urban areas in Africa are growing both in population and in their physical urban footprint; this growth is occurring in a context of weak institutions which have led to the prevalence of unplanned growth (Güneralp et al., 2017). The management of these cities requires constant monitoring of long term urbanization trends, to guide the construction of physical infrastructure but also its environmental and human impacts (e.g. the effect of urban growth on recurrent flooding).

This chapter summarizes the application of two methods operationalizing linear unmixing to generate long term land cover maps of urban systems based on mid-resolution imagery (Landsat images). The first method is a straightforward application of a supervised spectral mixture analysis (SMA), yielding results at sub-pixel level in the form of land cover fractions for each pixel. The second method consists of applying an unsupervised classification algorithm which makes use of SMA to generate land fraction maps, subsequently interpreted as land covers. The resulting classifications from these methods were assessed in terms of validation (comparison to independently derived land cover, from available high resolution imagery) and verification of the built-up land cover fraction.

The method was deployed to generate land cover maps for Kampala, Uganda (2001, 2005, 2010, 2016) and for Kigali, Rwanda (2000, 2009, 2015). The land cover maps derived in this chapter were used as inputs for further models throughout this dissertation.

2.2 Spectral mixture analysis: method and data

2.2.1 Spectral mixture analysis

SMA is a sub-pixel classification technique which considers the reflectance of any given pixel for any given band to be a function of the fraction of land cover and the land cover's reflectance value at this band, for a set of land cover classes. In particular, spectral linear unmixing assumes that, for a set of bands and land cover classes, the reflectance values of each pixel can be linearly decomposed:

$$x = \sum_{1}^{M} \mathbf{Sa} + \mathbf{w} \tag{2.1}$$

with $\mathbf{x}$ a vector of n spectral bands, $\mathbf{a}$ a vector with the fraction of M endmembers, $\mathbf{S}$ an $M \times n$ for which each entry is a reflectance value of an endmember for a specific band, each row represents an endmember's spectral signature (across all bands), and each column, the different endmember's reflectance at that wavelength, and $\mathbf{w}$ a vector of error terms.

SMA was implemented in R (R Core Team, 2017), using the library developed by Lehnert et al. (2016).

2.2.2 Data

For Kampala, Landsat images from row 171 and path 60 were acquired for 2001, 2005 (ETM+ sensor), 2010 (TM sensor), and 2016 (OLI sensor). Six bands (1-5 and 7) were selected from images sensed by the TM and ETM+, thus excluding the thermal band, which has a lower spatial resolution. To ensure comparability, six bands with equivalent wavelengths (bands 2-7) were selected from the 2016 image, sensed with the OLI.

For Kigali, three Landsat images, from path 172 and row 61, were acquired for the dates: July 12, 2015 (OLI/TIRS sensor); June 25, 2009 (TM sensor), and September 12, 2000 (ETM+ sensor). Six bands from each image (excluding the thermal band of images from sensors TM and ETM+, and choosing bands with similar wavelengths from the image sensed by the OLI to enhance comparability across time) were selected. For each band of all images, the pixel values were normalized to the range of 0.00 to 1.00.

2.2.3 Endmember selection strategies

Land cover in urban areas, derived from remote sensing, has been conceptualized over many years as composed of, mainly, three materials: impervious surface (buildings and pavements), vegetation, and bare soil (Ridd, 1995); a fourth element often included is water (Small, 2001), if large bodies of it exist within the extent being analyzed, and it is characterized by a very low albedo. These categories of land cover have been found to be lineally separable using SMA (Small, 2001; Kuang et al., 2014). Therefore, the conceptual model of land cover materials characterizes each pixel–a $900m^2$ square–as composed of four fractions, corresponding each to a land cover type and which sum 1.00. A subset of “pure” pixels show a fraction equal to 1.00 for one land cover type, and thus fractions of 0.00 for all other types. These pixels become key when applying the SMA, since they may be selected as endmembers in the process of linear unmixing.

To generate this classification of four land cover fraction maps, we operationalize the analysis by creating an auxiliary model of pixels with: (1) very high albedo (bright pixels in the visible spectrum), which may be either built-up structures or certain bare soils, (2) very low albedo (black features such as pavements or materials absorbing radiation such as water), (3) vegetation, and (4) bare soil. Since some features exhibit similar spectral signatures in the visible or near infrared wavelengths, it is possible to generate from the data several high albedo and low albedo groups (distinguishable in lower wavelengths). Hence the usefulness of this auxiliary model in matching auxiliary categories to land cover categories in a many-to-one manner.

The identification of pixels corresponding to endmembers was achieved using two different strategies, one for Kampala and another for Kigali; note that for these two methods, the result is a set of pixel locations:

For the Kampala case study, endmember samples were digitized by visual inspection of Landsat (combination of bands 4-3-2, the so-called false color, and of bands 7-5-4) and available high resolution imagery (of 2004, 2010, and circa 2016).

1. In first stage, four endmembers were selected. These were: high albedo (generally consisting of buildings, certain pavements, and soils), low albedo 1 (water, shadow, certain vegetation types such as wetlands), vegetation (typically shrubs, trees, and grassland), and soil (including most types of bare soils, although, as noted, certain smooth and very high albedo soils may be classed as high albedo).

2. To advance temporal consistency, the four endmembers for every year were sampled at the same locations: all selected locations corresponded to places that did not undergo land cover change.

3. Since these models ultimately caused a radical underestimation of built-up land cover fraction, a fifth endmember was added to the sample: low albedo 2 (dark buildings).
4. For this fifth endmember, it was not possible to sample the same location over the four years analyzed, because of the relatively rare occurrence of pure examples of these features within the city of Kampala.

For the Kigali case study, endmembers were selected using the unsupervised classification method proposed by Du et al. (2005):

1. The starting point (denoted m_0) is the pixel that is the maximum value of the sum of reflectance for all bands. This is a pixel with very high albedo.
2. Using only m_0 as the only spectral signature determining $\mathbf{S}$, a preliminary classification, $Class_0$, is performed.
3. From the resulting classification, m_1 is selected as the spectral signature from pixel with the largest error w_1, as estimated from equation 2.1.
4. m_0 is discarded. $Class_1$, the first classification of the method, is performed using only m_1 as spectral signature $\mathbf{S}$.
5. From the results of $Class_1$, the spectral signature of the pixel with the largest error (w_2) was extracted and denoted m_2.
6. $Class_2$, the second classification, was performed using both m_1 and m_2 as $\mathbf{S}$.
7. Steps 5 and 6 were repeated: for each classification performed, the pixel with the greatest error was extracted and added to $\mathbf{S}$; then, a new classification with the updated $\mathbf{S}$ was generated. When $\mathbf{S}$ reached six rows (equal to the maximum number of available bands for Landsat in a single year), $Class_6$ was generated.
8. The pixel corresponding to each row in $\mathbf{S}$ was extracted from the full database, including its coordinates (x, y).
9. The principal components of the normalized bands of all years (i.e., stacking up the three years into a single image of 18 bands and calculating the principal components of this image) were taken as input data, excluding only the pixels corresponding to cloud cover in the 2000 image.

Both methods were applied to each case study and the best results were obtained for each case by the selected method (supervised endmember selection for Kampala, unsupervised endmember selection for Kigali). Each method presents advantages and limitations: the endmember selection of the supervised approach need not be linearly separable, unlike the unsupervised approach. Thus, while the unsupervised results are guaranteed to be correct from a formal perspective, they will not necessarily correspond to the four categories of materials of the conceptual model. The supervised endmembers necessarily result all land cover categories being sampled, since they were purposefully selected with the conceptual model in mind. The unsupervised methodology does have the clear advantage of being fully replicable; supervised classifications are always subject to criticism, since many more potential endmember samples exist than those ultimately selected (although this objection is, to a point, answered by acceptable validation results).

2.2.4 Implementation of spectral mixture analysis

Since the result of the **Kigali SMA** were six categories but their corresponding classification (into high albedo, low albedo, vegetation, or soil) was not known, they were first sorted into one of these groups before being finally classified into land cover. The interpretation of endmember fractions proceeded as follows: the predicted fractions map corresponding to each endmember was reclassified into two, with the 95th percentile as the limit between them; patches of approximately 100 pixels of the category with larger fractions, recognizable at a scale of 1 : 50000, were compared to high resolution imagery when available (2008 and circa 2015) and to the original composites of Landsat imagery. Based on this comparison, each endmember was classed into: low albedo (easily recognizable due to the open water of the Nyabugogo River/Lake Muhazi in the north-eastern corner of the study area), high albedo (purely white in most Landsat composites), bare soil (yellow or light brown in the Landsat combination of bands 7-5-2), or vegetation (bright red in a Landsat false color combination).

The conversion of the auxiliary fractions into the final land cover fractions resulted from applying a rule-based approach. Categorical land cover maps were used as an indication of where potentially impervious surfaces may occur. **For Kampala**, maximum likelihood classification was used to sort cells into three categories: built-up, non-built, and water; built-up included buildings and pavements. The 2010 land cover map was adopted from Abebe (2013). To generate the 2016 map, a supervised classification was applied using Abebe's 2010 map as sample and the corresponding Landsat image as spectral data; similarly, Abebe's 1995 land cover map of Kampala was used to generate supervised classifications for 2001 and 2005 (with each year's corresponding Landsat image as spectral input). **For Kigali**, the GHSL of 2000 and 2014 (Pesaresi et al., 2016) were adopted as categorical maps of built-up land; for the year 2009, an unsupervised classification was created and the results, re-grouped into built-up, non-built, and water.

All categorical maps of **Kampala** were manually edited by visual inspection at a scale of approximately 1:50000. The objectives of this editing process were: (1) where and when (for the entire 2016 image and for the central part of Kampala, 2010 and 2004) high resolution imagery was available, to manually re-classify cells originally sorted into the built-up category but that were actually bare soil, (2) to ensure that large patches (e.g. 100 cells, i.e., visually evident at mid-resolution) that were classed as non-built in a later year were not classed as built-up in the immediately preceding year–the assumption here being that there was little to no large scale demolishing, and indeed when inspecting these cases, it was generally apparent that they conform to classification (methodological) errors rather than substantive phenomena. Since categorical maps of **Kigali** were based on a methodology already controling for these biases, no edit was performed for the Kigali maps.

Based on these categorical maps, the following rules were then implemented to regroup the auxiliary fractions into land cover fractions:

- All cells classified as water in the categorical classification are taken as having a water fraction of 1.00
- If a cell is built-up, according the categorical classification:
  - The sum of high albedo and low albedo fractions results in the impervious surface fraction (in the case of Kampala, only low albedo 2 was summed into the impervious surface fraction).
  - The auxiliary vegetation fraction equals the final vegetation land cover fraction (in the case of Kampala, to the vegetation fraction was added the low albedo 1 fraction); the auxiliary bare soil fraction equals the final bare soil fraction
- If a cell is not built-up, according the categorical classification:
  - The sum of high albedo and auxiliary bare soil fractions equals the final bare soil land cover fraction
  - The sum of low albedo (in the case of Kampala, of both low albedo 1 and low albedo 2) and auxiliary vegetation fractions equals the final vegetation land cover fraction

These rules seek to reflect the observable fact that high and low albedo fractions within the urban footprints of Kampala and Kigali (and generally of any city) are more likely to be impervious surface (a building with a white roof, asphaltic pavement or roof, etc.) than it is to be vegetation or bare soil. It does, however, introduce some degree of error, as high albedo smooth soils may be difficult to identify. The use of GHSL corrects, as much as possible, this bias, since it is a method that improves on traditional approaches in distinguishing bare soil from built-up (Pesaresi et al., 2016); further, the manual edits to the categorical maps also reduce this error.

The 2005 land cover map of **Kampala** exhibited a missing data problem because of the malfunction in the Scan Line Corrector of the ETM+ sensor in the Landsat 7 satellite. The gaps of the 2005 map were filled by a Krige extrapolation of impervious surface and vegetation fractions: a semivariogram was estimated based on the cells with data of 2005, it was fitted using an exponential model, and this model was used to calculate values for cells in the area with no data.

2.2.5 Validation of land cover maps

The validation strategy of the produced final land cover maps was adopted from Kuang et al. (2014). The impervious surface fraction was validated by:

1. reclassifying the map into 10 categories: ranges of $0.00 - 0.10$, $0.10 - 0.20, ...$, up to $0.90 - 1.00$;
2. within the area defined by each range, 10 points (at least $270m$, or nine cells, apart) were randomly sampled, although some ranges covered such a small area that the constraint prevented the algorithm from finding 10 points;
3. for each point, a square of 3×3 cells was digitized following the lattice defined by the land cover map and taking as the central cell that over which the point was overlayed;
4. the impervious surface fraction of each square was digitized from available high resolution imagery at a scale of $1 : 650$ approximately. For Kampala, this included aerial imagery provided by the city for 2004 and 2010, as well as satellite imagery from ArcGIS circa 2017; for Kigali, aerial imagery provided by the city for 2008 and satellite imagery from ArcGIS circa 2017;
5. these results were compared to the average impervious surface from the Landsat impervious surface fraction within each square, by calculating the root mean square error (RMSE) and the correlation between the data sets (the coefficient of regression for a linear model between them, with the high resolution imagery data as the dependent variable).

Note that the built-up land cover fraction includes pavements–it could perhaps be better described as impervious surface. In consequence, when digitizing the validation data, both buildings and paved areas (particularly roads) were included.

2.3 Land cover maps: results and discussion

2.3.1 General patterns and validation of impervious surface fractions

The generated patterns of built-up land cover fraction of both Kampala and Kigali are shown in figure 2.1.

Two simultaneous dynamics are apparent for Kampala in figure 2.1. Firstly, the city is undergoing a rapid outward expansion, a result broadly consistent with previous studies of urban growth of Kampala (Abebe, 2013; Vermeiren et al., 2012), much of it along the main road network, which has already surpassed the limits of the Kampala City Council Authority–the city proper's boundary. Secondly, a very significant amount of land use intensification has occurred, as witnessed by the increase of impervious surface in central locations (presumably a combination of more buildings for urban activities and new paved roads). Note that very rarely do impervious surface fractions over 0.80 appear in the maps–and the fractions of 2001 and 2005 are generally below the 0.40 threshold.

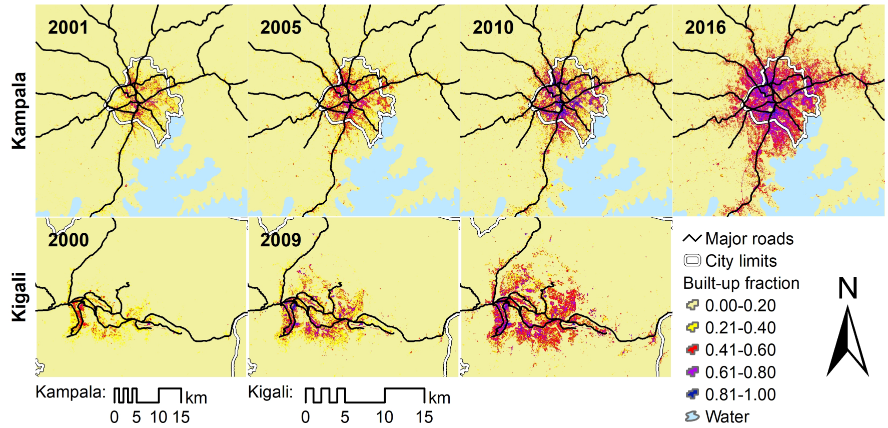


Figure 2.1 Built-up fractions for Kampala (2001-2016) and Kigali (2000-2015)

Kigali, as shown in figure 2.1, experienced a similar dynamic of outward expansion and intensification, which is predicted by the theory of urban location for cities undergoing very rapid population growth (Brueckner, 1987). However, Kigali is remarkable in that there is a change in the pattern of urban growth: For 2000-2009, greenfield development (especially towards the north of the city) resulted in the creation of new urban areas, essentially separate from the main urban patch that existed in 2000. In a second phase (2009-2015), both this new urban area and the older ones (i.e. existing in 2000) expanded outwards, which paradoxically resulted in the intensification of the urban fabric of Kigali. One may speculate two causes, not mutually exclusive, explain this dynamic. On the one hand, as noted by Goodfellow (2013a), around 2009 Kigali (and the Rwandan State) embarked upon an urban modernization project for the city highly reliant on the stringent implementation of a land use plan. On the other, Kigali presents a very irregular terrain, even more so than Kampala. Constraints due to excessive slopes may have partially caused the discontinuity.

The RMSE and correlation between the validation data and the land cover estimates is summarized in table 2.1. The sample size of the validation data is also reported. Recall most validation samples will have less than the 100 target points because areas of some built-up fraction ranges are so small that not enough points can be selected for sampling (and, since Kampala is a larger city than Kigali, the samples of Kampala are larger that those of Kigali precisely for this reason). Note also the validation data of Kampala, 2016 is larger than 100: this is because data collected for previous, discarded land cover models was kept for the final evaluation.

RMSE for Kampala (2005, 2010, and 2016) estimates vary between 0.218 and 0.237 and the correlation between the predicted value of the

Table 2.1 Comparison of impervious surface fraction results between land cover data set (Landsat based) and HRI data

City	LS model year	HRI data	RMSE	Corr.1/	n
Kampala	2016	2015	0.2371	0.7684	119
	2010	2010	0.2237	0.9069	93
	2005	2004	0.2177	0.8619	82
Kigali	2017	2015	0.1507	0.7406	64
	2008	2009	0.1590	0.7419	58

1/ Regression coefficient of LS data model on HRI data model

Landsat-based model and the high resolution sample was relatively high, over 0.768 for all years (table 2.1). These results are reasonably good, although the errors are higher than those reported for more industrialized contexts (see Kuang et al., 2014–their RMSE is below 0.175 and the correlation, higher than 0.89). This is likely a limitation posed by the phenomenon of urban land cover in Kampala and perhaps in other cities of Sub-Saharan Africa (though notably not Kigali), since there is a much larger fraction of bare soil relative to large cities of China and the US, Kuang et al.'s case studies. Like Kuang et al. (2014), the mid-resolution land cover model overestimates the impervious surface area fraction (even more so than their results), as witnessed by correlations that are less than 1.00 for all years. This limitation is also likely inherent and related to smooth, high albedo soil surfaces, the largest of which were manually edited out of the urban footprint maps but which likely remain, especially at sub-pixel level.

The RMSE and correlation between the high resolution land cover and the Landsat-based models for Kigali (table 2.1) are better than what was obtained for Kampala and similar to the errors reported by Kuang et al. (2014). Like Kampala, the results for Kigali show an overestimation of the impervious surface land cover, evident in correlations lower than 1.00. The results of Kigali show that the endmember spectra of this urban region are better at separating soil from impervious surface; possible explanations could include soil colors that are more distinct from built-up objects in Kigali (although Kigali is geologically more complex than Kampala), the use of different materials for building in these two cities, or an effect due to greater wetness of the soil and vegetation in Kampala, relative to Kigali. Alternatively, the unsupervised classification method does produce better results than supervised approaches (although the question would remain, in this case, why was the unsupervised algorithm incapable of improving on the results of Kampala too?)

2.3.2 A discussion on methodological choices and results

While ideally the product of this chapter would have been a unified method and results of comparable quality for both case studies, it is important to emphasize the priority is to ensure the best possible land cover data models (in terms of validation, i.e., correspondence to each case studies' land patterns). Indeed, given a choice between methodological uniformity and known limitations vs. reducing these limitations to the minimum, the latter road has been taken. This is because the land cover maps are inputs, crucial yet not central to the overall academic argument developed in this dissertation (in other words, the *crucial* aspect of them relates to their accuracy, not generalizability).

Several, unreported, land cover models were developed through the application of SMA for each case study. They differ in endmember selection and are variants of the two general approaches discussed in section 2.2. For each case, unsupervised endmember selection processes were deployed, seeking four, five, or six endmembers; these models used either transformed data of all available imagery jointly (as reported in section 2.2) or separately for each year. Similarly, several supervised endmember samples, of four endmembers (high and low albedo, vegetation, and soil) were generated for each city. From this relatively large pool of potential models, the final result was selected.

Two properties were used to judge each land cover model. The first, a verification issue, was consistency of urban patterns in time; the second, validation (comparing the land cover data model to independent data). Generally, making allowances for the additional error introduced by the presence of large bare soil areas, the validation property was acceptable (note, though, verification was first examined and used to discard land cover models; so it is likely the less successful models, verification-wise, would have been equally unsuccessful in terms of validation). However, the verification aspect often represented an insurmountable obstacle to accepting a land cover data model.

The fundamental problem of verification is the following: both case studies represent cities undergoing rapid population growth and, consequently, a physical expansion of the built-up area. On aggregate terms, this means that if the built-up land cover fraction map of a later year were to be subtracted to an earlier year, one should expect the overwhelming majority of cells to be greater than 0.00 (negative values correspond to either classification errors or, more rarely, demolitions). Further, the total area of negative cells (the summation of negative cell values times the cell area) should be very small, since they are errors.

Table 2.2 shows the breakdown for such map algebra operations (which amount to maps of urban growth) for each pair of succeeding years. As can be seen, the internal consistency of the selected built-up land cover fraction maps is very good and the overall error, small. In most alternative (discarded) versions of the land cover data models, the total negative areas represented up to 40% or more of the total change (note that, while the percentages on table 2.2 may seem large, they are estimated relative to the change, not the total cells as is usual; no change cells make up over 80% of all cells); further, when examining the pattern, it was evident the cause of this large error was systematic under- or overestimation of the built-up fraction for a given year (this was especially true when the urban core around the CBD was the area

Table 2.2 Urban growth: resulting areas per built-up fraction (ha)

		Built-up fraction range		
City	Period	< -0.01	$[-0.01, 0.01]$	> 0.01
Kampala	2001-2005	1124.8	1.1	4071.5
		21.6%	0.0%	78.3%
	2005-2010	1567.3	0.8	5458.7
		22.3%	0.0%	77.7%
	2010-2016	1461.3	0.3	9729.2
		13.1%	0.0%	86.9%
Kigali	2000-2009	911.4	0.6	13410.0
		6.4%	0.0%	93.6%
	2009-2015	133.8	0.1	1608.5
		7.7%	0.0%	92.3%

Percentages relative to sum of change area (row).

showing a concentration of negative cells).

2.4 Synthesis

The land cover fraction maps for four categories (built-up, vegetation, soil, and water) were presented and discussed. The methodologies through which they were generated were documented. The validation and verification of the resulting models were summarized; they were discussed in the context of (unreported) discarded land cover models.

In general, both cities were found to expand both outwards from urban centralities (greenfield development) and by intensification of central locations. This pattern is consistent with theoretical consequences of exogenous population growth (Brueckner, 1987). The expansion of Kigali was also affected by the implementation of the land use planning system circa 2009 (see OZ Architecture et al., 2007 and SURBANA International Consultants PTE Ltd., 2013, as well as the results of chapter 6), which encouraged more compact development.

Land cover data models of Kampala and of Kigali show consistency in time (relatively few cells exhibit larger built-up fractions for earlier years than for later years), a property used to verify an observed characteristic of each urban system, i.e. that they are experiencing evident urban growth. Land cover maps were also found to be acceptable in terms of validation (by comparing them to independent data derived from high resolution imagery).

In sum, the land cover map results reported in this chapter were judged to be accurate and, thus, an acceptable input for further modeling developed in this dissertation.

2.5 Appendix

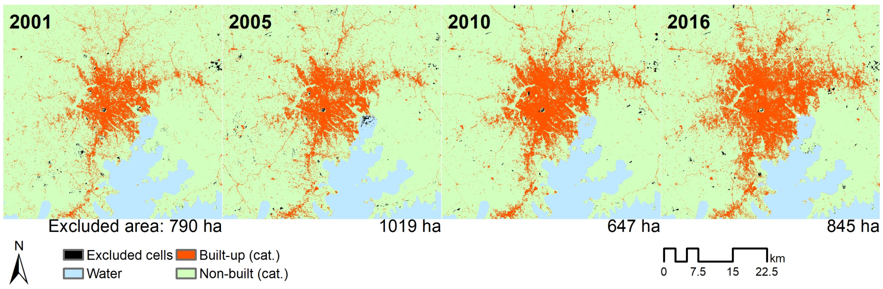


Figure 2A.1 pixels excluded through manual editing process in Kampala

3

Structural equations modeling of the impact of flooding on urban patterns[1]

Abstract

Recent developments in socio-hydrology, as well as microeconomic models of urban location, suggest exposure to natural hazards may constrain urban development. Cities in Sub-Saharan Africa often suffer recurrent flooding and weak institutional settings which preclude land use planning systems becoming effective risk mitigation tools. How does human behavior on location choice factor into this context? A structural equations model was developed to understand the causal paths between population growth, hydrological impacts, and urban growth, using data from eight catchments of the Kampala metropolitan region over three periods (2001-2005, 2005-2010, and 2010-2016). Urban growth is conceived as being triggered by exogenous population growth which also increases flooding via the larger impervious urban areas it causes, re-distributing urban growth away from subcatchments with larger hydrological impacts. The structural equations models are generally found to fit the data, providing evidence to confirm the hypothesized causal paths. Marginal effects of peak flows on urban growth were found to be statistically significant, although small (0.2 standard deviations from the mean). Direct effects of population growth on urban growth were much larger (between 0.6 and 3.0 standard deviations from the mean, depending on the causal path). The constraining effect of recurrent flooding is, thus, not large enough to mitigate flood risk by itself, but it does open a window of opportunity for land use planning in the face of rapid population and urban growth.

Keywords: land cover change, flood impact, path analysis, structural equations modeling, Kampala (Uganda)

[1] This chapter is based on: Pérez-Molina et al. (2019a).

3.1 Introduction

Kampala, Uganda is a city of over two million residents (United Nations, 2018) which undergoes widespread and recurrent flooding due to relatively small rainfall events and low infiltration soils (Sliuzas et al., 2013). Such flooding results in impacts on the quality of life of urban residents, particularly the poorest (Isunju et al., 2016), as well as repeated disruption of city functions, especially transportation (Lwasa, 2010).

During the last twenty years, the city of Kamapala has embarked, with the World Bank, in a program of systematic investment on road and drainage infrastructure (World Bank, 2013). Specifically, the main drainage channels of the city have been designed and are being expanded, to increase their capacity to carry runoff and thus reduce impacts of flooding on city functions. The rate of progress in improving drainage has been slow; work has concentrated on two out of the main eight primary drainage channels (Nakivubo and Lubigi). Thus, flooding is expected to remain a problem, even after the infrastructure investment program has been completed.

In this context, two research questions can be posed, to clarify the impact of urban growth on flooding: (1) To what extent are urban agent's location choices influenced by flooding? (2) How does this influence manifest itself, if at all, on aggregate urban patterns? The hypothesis is that flooding acts as a constraint on urban development; but the aggregate effect of this constraint should be limited, since many urban agents who otherwise would be excluded from desirable locations are likely willing to accept the risk of natural hazards if compensated through greater accessibility or amenities.

These issues were addressed by deriving a causal path analysis model, based on microeconomic theory and the socio-hydrology framework. In this way, relevant factors were organized to reflect the relations expected from the theory. Structural equations models are used to test whether the data assembled to describe Kampala is consistent with the hypothesized causal paths. The results are discussed in view of the need to markedly improve the capabilities of land use planning systems, especially with regard to regulation, in Kampala.

3.2 Conceptual framework and previous work

3.2.1 Socio-hydrology and feedback mechanisms between human settlement patterns and flooding

During the last decade, the advent of socio-hydrology (Sivapalan et al., 2012) has promoted (mainly from the hydrological viewpoint, see Xu et al., 2018) the exploration of relations between hydrological and human systems. Conceptualizations of human settlement patterns and flooding, particularly those considering the well known levee effect, have provided important insights on the analysis of dynamic interactions and feedbacks, conceiving both hydrological and social subsystems as linked processes. Notable among them is the effort of Di Baldassarre et al. (2013), who simulate the interplay between economics (overall wealth and population size), technology (levees), society (awareness of potential risk), and urban development patterns (affected by the memory of past floods).

While conceptual improvements have contributed to clarify the issues and to provide mathematical frameworks for exploring specific cases, the development of evidence by means of case studies has been limited. Hall et al. (2014) characterized these low dimensional models as indicative and complementary to the more complex (and data-demanding) process-based or comparative case study analysis. After searching the literature for empirical case studies, three (Odongo et al., 2014; Chen et al., 2016; Zischg et al., 2018) have been chosen because they provide insights into the methods to derive substantive conclusions on socio-hydrologic processes.

Zischg et al. (2018) adopted a backcasting approach. They modeled river morphology (before and after the construction of levees and the deepening of river channels), land use patterns (based on current and historical buildings, as well as with alternative zoning scenarios, i.e. eliminating buildings according to constraint maps), and an inundation model plus vulnerability functions to estimate damage (by assigning flood depths to the buildings identified as part of the land use pattern) for the Emme River in Switzerland. Based on these elements, they evaluated possible alternative paths by simulating the 1910-2015 period and tracing the evolution of flood costs.

Chen et al. (2016) also simulated a relatively long time series (1948-2015) for the Kissimmee catchment, using a model similar to Di Baldassarre et al. (2013); this catchment was first channelized to mitigate flooding (during the 1960s) and it was later restored to promote wetland health (in the 1990s). Chen et al. (2016) simulated these management strategies as the product of a change in values and in relative power of upstream vs. downstream population, although informed by flood damage (i.e. as flood control measures were successful, society placed a decreasing importance on flood in determining policy). A feedback effect was thus introduced between flood, values, and policy, with land use consequences in terms of total wetland area.

Odongo et al. (2014) identified causal cascades and quantified them, with limited data, using path analysis for the Lake Naivasha land use/water abstraction system. Causal cascades were triggered by economic development evidenced in flower production area (upstream) and in population growth (downstream), leading to ecological and hydrological impacts. This case is, methodologically, the most similar to the results reported in this chapter, as it applies path analysis, it is also located in Sub-Saharan Africa, and makes use of limited data (the latter two conditions being linked and explaining the choice of method) to tackle a feedback between human land use and its hydrological impacts.

In the three cases (Odongo et al., 2014; Chen et al., 2016; Zischg et al., 2018), the feedback loop was decomposed into a linear causal path, making the problem tractable from a quantitative perspective. The theoretically hypothesized feedback effect was incorporated (by Chen et al., 2016 and by Odongo et al., 2014) through the dynamic nature of the system: events from the preceding period affect the succeeding ones. This is consistent with Shipley's (2016, pp. 36–38) argument that feedbacks in a causal process are a consequence of not explicitly accounting for time. The second central element that should be highlighted is the spatial unit: all three studies aggregate their results into (sub)catchments; Chen et al. (2016) and Odongo et al. (2014) separate their own into upstream and downstream subcatchments, in recognition of the diversity of social and physical processes occurring there. Finally, available data was constrained: while Chen et al. (2016) deploys a full quantitative dynamic analysis, both Odongo et al. (2014) and, more transparently, Zischg et al. (2018) use limited time steps in their models, reflecting the evolution of the system in time not as a strictly dynamic model but rather as a succession of static states. Yet this is sufficient to detect and characterize the relations between human and natural systems.

3.2.2 Urban land cover patterns and location choice under natural hazards

The basic microeconomic model of urban location explain the choice of urban agents as a trade-off between commuting costs and housing prices (Brueckner, 1987). From the model, one may derive that the capital-to-land ratio is a decreasing function and yard space, an increasing function (Brueckner, 1983) of commuting costs; these conclusions, taken together, mean the built-up land cover fraction is a decreasing function of distance to the CBD. Additionally, an exogenous population increase should lead to a physically larger city (its urban footprint increases) with smaller dwellings and larger capital-to-land ratios in all locations (the city densifies, see Brueckner, 1987).

Frame (1998) extended the basic model to account for the exposure of land to flooding by introducing a loss caused by the flood, which subtracts from the urban agent's income. This loss affects only exposed locations and it is also a function of the event's severity (large floods cause greater losses than small floods). One must note that this is similar but not equivalent to cumulative losses from very small events occurring very often; the difference lies in the need, for a cumulative approach, of a dynamic system (i.e. the need to consider successive periods), which is difficult to formalize within a static neoclassical model. From Frame (1998), one may conclude that capital-to-land ratio is a negative function of potential risk, which is to say there is less building in areas of the city exposed to flood. Moreover, one can demonstrate the existence of a critical level of potential risk; for areas within the city, if exposed to greater potential risk than the critical threshold, no urban development takes place. The critical level of potential risk is a negative function of commuting costs: households are compensated for greater potential risk (and the losses it entails) by less commuting

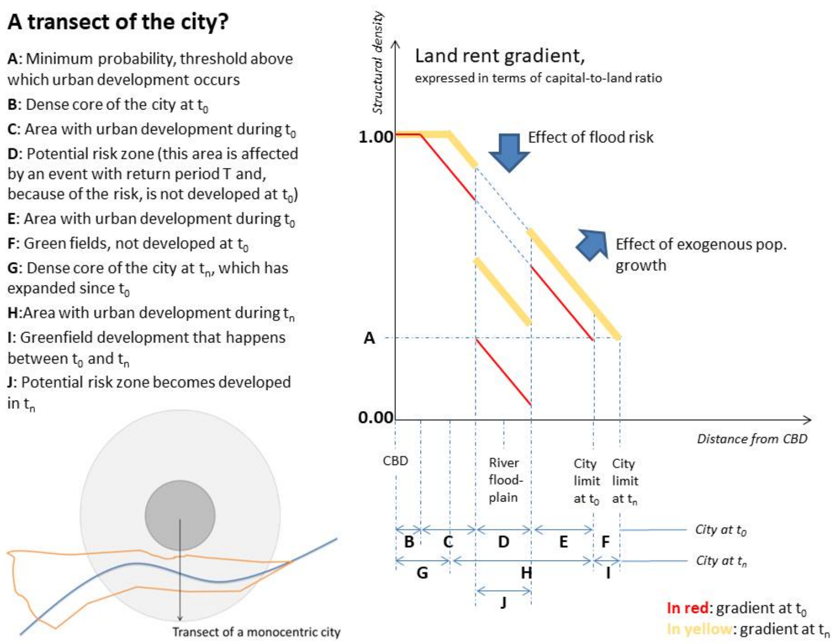


Figure 3.1 Stylized urban land gradient derived from Frame (1998). Bottom left: a monocentric city, with the center in dark grey and the periphery in light grey; the blue line represents the river and the red line encloses the area affected by the flood event. Right: the transect, showing the variation of strutural density as distance from the center increases.

time. One should, again, note that the assumption of a monocentric city has carried on from the Alonso's original model; however, it can be relaxed to incorporate job decentralization–the observable fact, in Kampala as in a large variety of urban contexts, that job locations are dispersed throughout the metropolitan area (Glaeser, 2008). Finally, given an exogenous population increase, the critical risk level increases (the undeveloped area within the city becomes smaller).

The physical consequences of the basic model are shown, schematically, in figure 3.1: the y-axis represents the probability of a plot being built-up and the x-axis, the distance from the CBD. As population in a city increases, the probability of all locations becoming built-up increases (due to larger structural densities and to dwellings becoming smaller, represented by the shift towards the right, i.e. away from the center, of the urban gradient) and the urban footprint of the city also becomes larger (at time period t_n, the built-up location furthest from the CBD, area **I**, is at a greater distance than the corresponding location for t_0, area **E**).

The effects of potential flood risk are also schematized in figure 3.1: the inner boundary shown as **D** for t_0 and **J** for t_n. Since from t_0 to t_n a population increase has occurred, the inner boundary disappears: the area **D** was not developed whereas the area **J** is, albeit at a lower probability than less accessible but not exposed to potential risk locations (corresponding to the area **E**, which at t_n should exhibit larger structural density, since it is not exposed to potential flood risk).

3.2.3 Proposed path models of urban growth and flooding

Causal path models are advanced, incorporating the conceptual relationships described in subsections 3.2.1 and 3.2.2, as well as a conceptual understanding of flooding within a rainfall-runoff modeling approach: rainfall falls over the landscape, part of the water is infiltrated into the soil and part flows over the landscape, eventually reaching drainage channels; there, it may overflow, if the capacity of the channel is exceeded.

The proposed causal paths are shown in figure 3.2; variable definitions and data sources are discussed in subsection 3.3.1. All causal paths are triggered by an exogenous population growth (*Pop Growth*), which should lead to larger urban areas (*Urban Area*) and urban growth (*Urban Growth*); note that larger urban areas do not necessarily correspond to bigger urban growth: for example, a spatial unit may be completely occupied by built-up land cover (large urban area with little possibility of urban growth). The central section of the causal paths reproduces the relations from the rainfall-runoff approach: greater urban areas mean more impervious area (less infiltration), ultimately resulting in greater peak flows (*Peak Flow*); larger rainfall events (*Rain*) imply more water eventually reaching the surface, likewise leading to greater peak flows. These basic physical relations are captured in the causal path corresponding to figure 3.2a.

The second causal path, figure 3.2b, controls for accessibility by introducing the *Central* variable: a dichotomic variable equal to 1.00 for the upstream part of catchments located around the CBD, as well as the downstream part of the Nakivubo basin (which is part of the CBD itself). This centrality variable should cause greater urban areas but its effect on urban growth is ambiguous: while central locations are more attractive for urban development, they are also older areas of the city–and therefore, locations where little space is available for new urban development.

The most complex causal path, figure 3.2c, introduces *location choice* as a latent variable–an unobserved effect that can be inferred. Such variable was hypothesized to replicate (at an aggregate level) the methodological approach of spatially explicit models, such as cellular automata, dependent on a single rule (e.g. a transition rule) to find the locations that are first chosen by urban agents. It also presents the advantage of separating the spatial structure (with relatively little variation in time) from the dynamic trigger of urban growth.

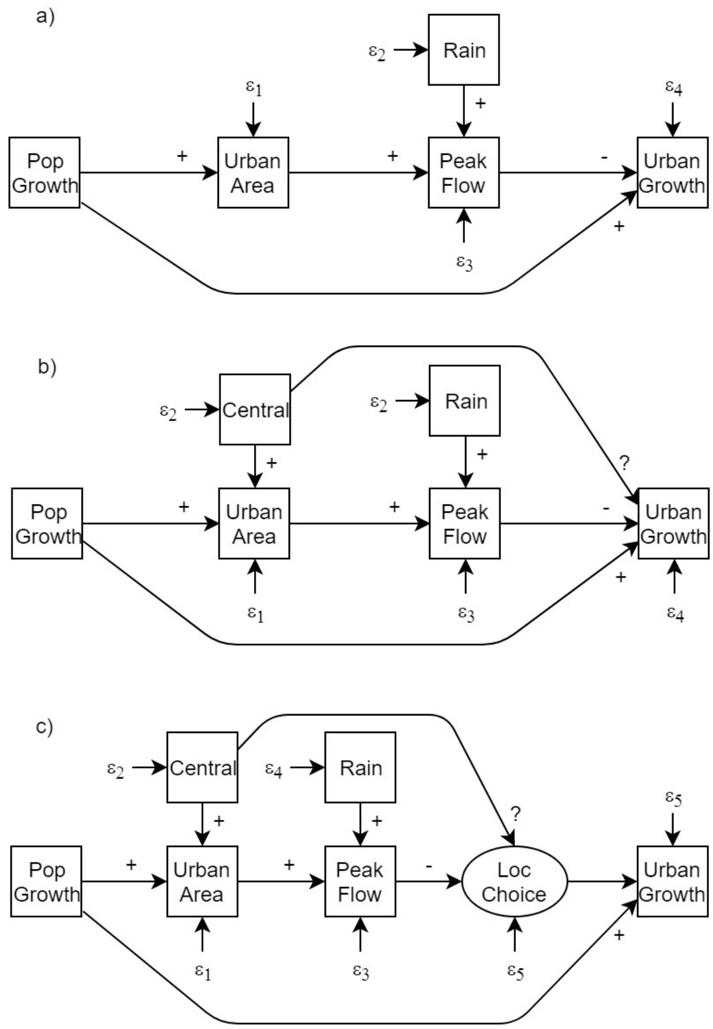


Figure 3.2 Proposed causal paths. a) Reduced form causal path incorporating main physical relations derived from location choice modeling. b) Causal path controlling for accessibility (centrality) and physical factors. c) Causal path incorporating location choice as a latent variable and controlling for accessibility effects.

3.3 Structural equations modeling methodology

3.3.1 Study area and data

The metropolitan region of Kampala is formed by the city proper, the boundary over which the Kampala Capital City Authority (KCCA) exercises jurisdiction, and the surrounding areas of the Wakiso district (see figure 3.3). Kamapala is located on the northern shore of Lake Victoria,

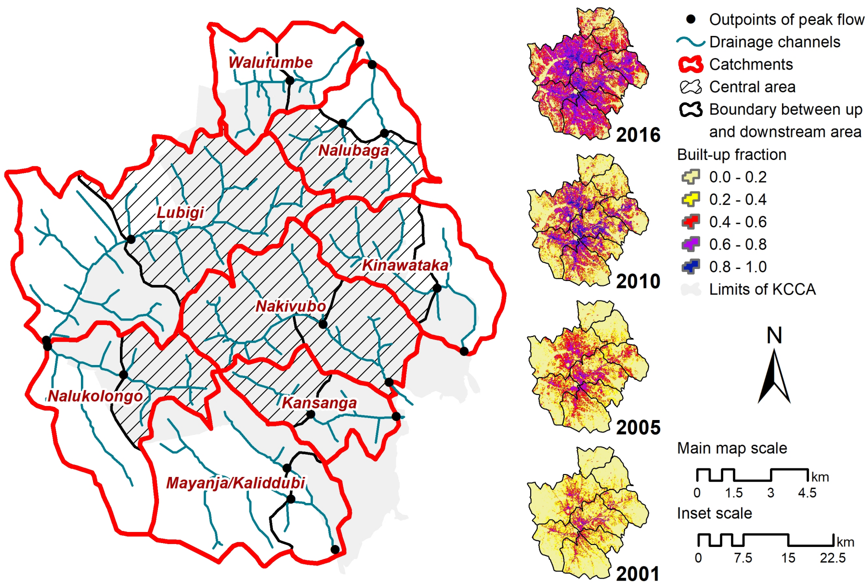


Figure 3.3 Study area: catchments and their subdivisions, built-up land cover fraction (2001-2016), and location of central areas for Kampala, Uganda.

a hilly terrain with valley floors occupied (under natural conditions) by wetlands. As the city has expanded, the wetland areas have been constantly encroached upon, increasing in this way the exposure of human settlements to flood risk (Isunju et al., 2016).

Given its physical context and the theoretical evidence on urban patterns, Kampala has been judged to be a representative case study to test the impact of hydrological outcomes on urban patterns. The analysis was operationalized by compiling data on population, built-up land cover, rainfall, and flooding. To perform hydrological analysis, the drainage catchments defined in the Kampala Physical Development Plan (ROM Transportation Engineering Ltd. et al., 2012), in turn adopted from the Drainage Master Plan, were selected (two small catchments in the periphery, Mayanja North and Gaba, were excluded because hardly any flooding or urban growth occur in them). All catchments were extended beyond the KCCA using a digital elevation model provided by the KCCA. They are shown in figure 3.3.

Each catchment was subdivided into two parts, upstream and downstream. The boundary between these parts follows the greatest slope and passes through the middle point of each main drainage channel. This distinction (between upstream and downstream) was introduced to control for variations in both hydrological and urban dynamics: generally, upstream areas have been occupied by urban land uses for longer and the drainage system has been channelized, downstream areas present better conserved wetlands (in general) with natural drainage and less human settlement (ROM Transportation Engineering Ltd. et al., 2012).

Since the objective of this chapter is to relate the outcome of hydrological processes, records in the data set are each of the upstream or downstream area of every catchment. As there are eight catchments in the study area, each divided in two, the cross section of the data is composed of 16 geographical units. Data is available for the years 2005, 2010, and 2016; the final data set has, in consequence, 48 records. It is included as part of the appendix to this chapter.

Population estimates for Kampala

Population for each hydrological unit was estimated by compiling census data for the five sub-districts of the KCCA (shown in table 3.1). For each intercensal period, yearly growth rates were calculated per sub-district. The census data was adopted as base year (1991, 2002, 2014) and, with the corresponding growth rater, was used to project population to the target years of 2001, 2005, 2010, and 2016.

Table 3.1 Population projections for Kampala sub-districts

	Census data			Yearly growth rate	
Sub-district	1991	2002	2014	1991-2002	2002-2014
Kampala	112787	88094	75168	−2.22%	−1.31%
Kawempe	158610	262165	338665	4.67%	2.16%
Makindye	186997	303171	393008	4.49%	2.19%
Nakawa	136519	240624	317023	5.29%	2.32%
Rubaga	179328	295088	383216	4.63%	2.20%
	Population projection				
	2001	2005	2010	2016	
Kampala	90095	84668	79251	73206	
Kawempe	250458	279495	310960	353430	
Makindye	290142	323494	360438	410381	
Nakawa	228540	257796	289183	331932	
Rubaga	282025	315010	351247	400275	

The built-up land cover maps were used to calculate the total built-up area for each sub-district of the KCCA. Gross population density was estimated by dividing the total built-up into the estimated population for each target year. Maps of gross density were created by assigning the estimated density to each urban pixel (defined as any pixel with built-up land cover fraction greater than 0.00).

The final population estimate per hydrological unit was achieved by: calculating the average (of all urban pixels) gross density per catchment and multiplying this gross density times the total built-up area of the hydrological unit. Population growth was calculated by subtracting the target year population minus the preceding period's total population estimate. Such estimates form the data corresponding to the variable

Pop Growth of the proposed causal path models (figure 3.2). The *Pop Growth* of 2001-2005 was assigned to the year 2005, of 2005-2010 to the year 2010, and of 2010-2016 to the year 2016 (the same procedure was used to assign urban growth data to the different years).

Land cover data models of Kampala

The development of land cover maps for the metropolitan region of Kampala, based on Landsat imagery and sub-pixel classification methods, was described in the chapter 2.

Built-up land cover fractions were aggregated to estimate total built-up areas by summing all pixels within a hydrological unit and multiplying this value times the cell size. These estimates of total built-up area correspond to the variable *Urban Area*; the variable *Urban Growth* was estimated by subtracting the target year total minus the preceding period.

Rainfall data model of Kampala

Rainfall data correspond to maximum daily total per year; daily totals were compiled for 2001-2016. Rainfall estimates were taken from the Rainfall Estimator (RFE2.0) for Africa (Love et al., 2004), available on-line from the IRI/LDEO Climate Data Library (International Research Institute for Climate and Society, Columbia University, 2018). Data was queried for coordinates $0°18'49''N$, $32°34'52''E$, corresponding to Kampala.

Rainfall events were simulated based on the selected daily total ($78.3mm$ for 2005, $165.7mm$ for 2010, and $89.1mm$ for 2016) and the procedure developed for East Africa by Fiddes and Forsgate (1974), which is based on the Intensity-Duration-Frequency approach. Specifically, for Kampala:

$$R_t = \frac{t}{24} \cdot \left(\frac{24 + 0.33}{0.33 + t}\right)^{0.95} \cdot R_{d,T} \tag{3.1}$$

with R_t the rainfall depth for time t (in hours) and $R_{d,T}$ the corresponding $24h$ total.

First, the total rainfall was estimated for a $15min$ period; then, the total rainfall corresponding to $45min$ was estimated; as a third step, the total rainfall for the $15min$ immediately preceding (and succeeding) the peak rain was calculated as $(R_{45min} - R_{15min})/2$. The process was repeated for every $15min$, until a $165min$ event had been obtained (and, as a final correction, the total for $60min$ was calculated to include the $30min$ preceding and succeeding this rainfall event). Thus, a $225min$ rainfall event was obtained for each year (these hyetographs are included in the appendix).

Flood data model of Kampala

Peak flow estimates were generated using the *OpenLISEM* flood model (Jetten, 2018): a rainfall-runoff flood model designed for mid spatial resolution (cells of $5m$ to $50m$) and very detailed temporal resolution (0.1 to 60 seconds). The flood model for the Kampala metropolitan region was developed by Umer et al. (2019) at a $20m$ resolution. A previous model of *OpenLISEM* was a central component of the IFM Kampala project (see Sliuzas et al., 2013); it concentrated on the Upper Lubigi sub-catchment (roughly coinciding with the upstream Lubigi hydrological unit of this chapter).

This model was re-sampled to a $50m$ resolution and then employed to simulate three rainfall events. Each of these scenarios was specified with: the corresponding rainfall event described in the preceding subsection and the land cover model (of built-up, vegetation, and soil fractions) of the years 2005, 2010, and 2016. Since the rainfall event is the maximum daily rainfall event of the five-year period *preceding* the target year, an implicit assumption is that the land cover of the target year is a reasonable model for when the peak rainfall event took place. Furthermore, the models generally reported in this chapter assume urban location choice is informed by a relatively large event in the recent past, rather than smaller, more recent events or larger catastrophic events of the past.

Peak flow estimates were produced for the outpoints shown in figure 3.3. Notice some upstream hydrological units (Mayanja/Kaliddubi, Nalubaga) have two outpoints along the main drainage system: peak flows of each were summed, although this correction was minor, since in both cases most of the flow passed through one of the two channels in question.

3.3.2 Structural equation models

The models proposed in figure 3.2 embody a number of hypotheses expressed as relations between variables and formalized through parameters. Structural equations modeling were developed to test the validity of such a causal arrangement. The process is described in Shipley (2016, p. 89): (1) *first*, a causal path analysis must be specified; (2) this causal structure is then translated into a set of equations (parameters in these equations–slopes, variances, covariances–may be estimated from the data or can be fixed as part of the causal hypothesis); (3) for each pair of variables, the variance and covariance are calculated from the data; (4) the maximum likelihood method is applied to determine the parameters that are not fixed, minimizing the observed covariances (from the data) and the covariances predicted by the model; (5) the probability of having observed the measured minimum difference between covariances from the data and from the model prediction is calculated (this is the null hypothesis, that the difference between observed and predicted is due only to random sampling variation); (6) if the calculated probability is sufficiently small, the proposed model is rejected.

Note that structural equations modeling is a *confirmatory* analysis (Shipley and Lechowicz, 2000). If the null hypothesis is true, the maximum likelihood χ^2 statistic follows a χ^2 distribution, with the degrees of freedom dependent on the model specification and the data sample size.

SEM analysis does not have, as an objective, to reject the null hypothesis; rather, since there are strong theoretical reasons to opt for the causal model, strong statistical evidence (i.e. a very low probability of the χ^2 statistic) is required to reject the model. Else, the model is judged to be a good fit to the data and to provide evidence to confirm the hypothesis.

Latent variables are variables that form part of the hypothesized causal path and that cannot be directly observed. Information about these variables can be obtained from other variables causally linked to them (Shipley, 2016, p. 127). Since the latent variables have no units in themselves, as they cannot be observed, it is a common practice to fix a path coefficient of one of the arrows leading into the latent variable to 1.00; thus, the latent variable has the same scale as the causal parent of the corresponding fixed path coefficient; alternatively, the variance of the latent variable may also be fixed to 1.00 (Shipley, 2016, p. 141).

Due to the reliance of structural equations modeling on maximum likelihood, limitations on their application due to characteristics of the data must be addressed to validly apply the method: (1) inferential tests are asymptotic, requiring potentially large sample sizes, (2) functional relationships must be linear, (3) data must be multivariate normal (Shipley, 2016, p. 61).

All variables were transformed into their natural logarithms to improve normality. A scatterplot matrix of the (transformed) data is shown in figure 3.4; most relationships are acceptably linear, with the exception of *Rain* and *Central* (these can be thought of as ordered categorical variables). In addition to the transformation, the Satorra-Bentler correction (Savalei, 2014) was used to produce a χ^2 estimate and probability that are robust to non-normality.

Regarding sample size, a rule of thumb requires at least 5 records for each parameter estimated (Shipley, 2016, p. 167). Since the sample consists of 48 records, it is perhaps barely enough to estimate all the parameters involved in the causal path models hypothesized in figure 3.2. Therefore, in addition to estimating with the sample data set, the Bolen-Stine bootstrapping method was applied, using the sample data-based model to predict 5000 data sets, each of which imposes the covariance structure model on the data (Kim and Millsap, 2014). For each data set, the structural equations model was fit and the robust χ^2 was estimated. In section 3.4, Shipley and Lechowicz (2000) were followed by reporting the 95% percentile of the Monte Carlo-based χ^2 probability for each model, in square brackets, next to the sample data-based probability (because of severe convergence problems, the model with the latent variable only produced 3288 valid results from the 5000 iterations; for that model, the χ^2 was reported on this smaller bootstrapped simulation).

Structural equations models were fitted using the *lavaan* package (Rosseel, 2012) from R (R Core Team, 2017): the *sem* function was used to fit the models using the sample data and the *lavaanBootstrap* function, for bootstrapping with the Bollen-Stine bootstrapping method.

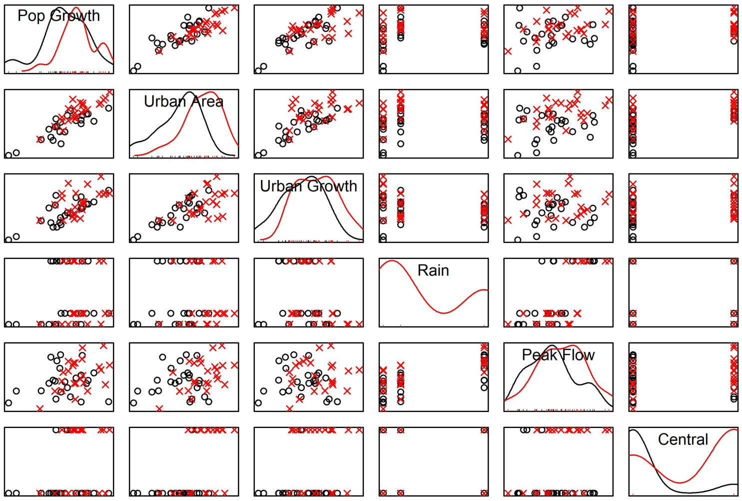


Figure 3.4 Scatterplot matrix of input data. Black circles denote downstream units, red crosses the upstream units. Data shown as natural logarithm transformation (non-transformed values reported in supplementary material).

3.4 Results

The reduced form path model (figure 3.2a) did not show evidence of lack of fit ($\chi^2 = 2.425$, $df = 4$, $p = 0.658$ [0.335]). Standardized path coefficients for this model, expressed in standard deviations from the mean, are reported in figure 3.5a: as can be seen, they are all significant (but for the path coefficient of *Peak Flow* on *Urban Growth*) at a $p = 0.01$ or better; the probability of each path coefficient can be seen in equation 3.2. Additionally, the signs of all path coefficients (figure 3.5a) coincide with the hypothesized effect (in figure 3.2a). As perhaps should be expected, the weakest effect in terms of significance and of marginal effect (absolute value of the path coefficient) is that of flooding on urban growth, which suggests accessibility and other constraining factors (especially income of urban agents, not considered because of the aggregate level of the model and this is an individual-level effect) are more important than physical constraints to the urban location choice underlying the urban patterns. However, it is also important to note the result does confirm the main hypothesis, namely that hydrological impacts of natural systems act as a constraint on urban growth; furthermore, this effect can be detected at an aggregate (subcatchment-scale) level with spatial data.

$$
\begin{array}{llllll}
\mathit{UrbanArea} & = & 0.927 \cdot \mathit{PopGrowth} & & & \\
 & & (< 0.001) & & & \\
\mathit{PeakFlow} & = & 0.254 \cdot \mathit{UrbanArea} & + & 1.873 \cdot \mathit{Rain} & \\
 & & (< 0.001) & & (< 0.001) & \\
\mathit{UrbanGrowth} & = & 0.672 \cdot \mathit{PopGrowth} & - & 0.172 \cdot \mathit{PeakFlow} & \\
 & & (0.002) & & (0.090) &
\end{array}
\tag{3.2}
$$

The second structural equations model, corresponding to the second hypothesized causal path (figure 3.2b), also shows a good fit to the data ($\chi^2 = 1.241$, $df = 5$, $p = 0.941$ [0.416]). With the exception of two path coefficients (that of *Peak Flow* on *Urban Growth*, significant at $p = 0.10$, and that of *Central* on *Urban Growth*, not statistically significant even at $p = 0.10$), all path coefficients are significant at $p = 0.01$ or better (the probability associated to each path coefficient can be seen in equation 3.3). Like in the structural equations model of figure 3.5a, in this second model (figure 3.5b) the marginal effect of the hydrological impact is also the weakest, as is nearly its significance. Yet it also provides evidence confirming the main hypothesis (the constraint of flood risk on urban growth).

$$
\begin{array}{llllll}
\mathit{UrbanArea} & = & 0.834 \cdot \mathit{PopGrowth} & + & 0.865 \cdot \mathit{Central} & \\
 & & (0.001) & & (< 0.001) & \\
\mathit{PeakFlow} & = & 0.254 \cdot \mathit{UrbanArea} & + & 1.873 \cdot \mathit{Rain} & \\
 & & (< 0.001) & & (< 0.001) & \\
\mathit{UrbanGrowth} & = & 0.645 \cdot \mathit{PopGrowth} & - & 0.203 \cdot \mathit{PeakFlow} & \\
 & & (0.003) & & (0.051) & \\
 & + & 0.331 \cdot \mathit{Central} & & & \\
 & & (0.184) & & &
\end{array}
\tag{3.3}
$$

A useful feature of structural equation models is the possibility to decompose causal effects along different paths (Shipley, 2016, pp. 105–109). Particularly interesting is the effect of the trigger of the urban growth process, population growth (represented by the *Pop Growth* in the causal path models proposed in figure 3.2) on urban growth. There are two such causal effects: the direct effect, represented by the path coefficient of the arrow directly linking *Pop Growth* to *Urban Growth* in figure 3.5b. Because population growth causes a larger urban area, which in turn produces greater peak flows, there is also an indirect effect: it is the result of multiplying the path coefficients of the arrows linking *Pop Growth* to *Urban Area*, *Urban Area* to *Peak Flow*, and *Peak Flow* to *Urban Growth*.

The direct effect of population growth on urban growth, as can be read off the causal model reported in figure 3.5b, is 2.963; the indirect effect is equal to -0.040, two orders of magnitude less. The indirect effect of population growth on peak flows, i.e. on hydrological effects, is 0.2119: about five times that of the indirect effect on urban growth. These effects are roughly consistent with their equivalents of the first causal

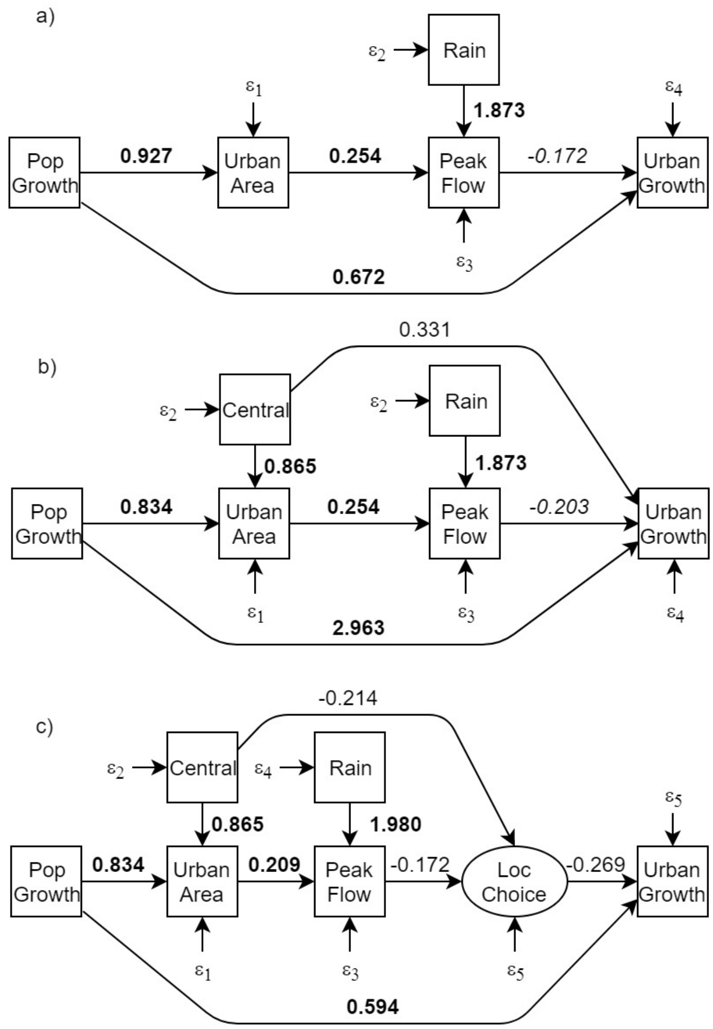


Figure 3.5 Result of path analysis for causal paths of figure 3.2: standardized path coefficients reported for each causal link. Units of the coefficients are standard deviations from the mean. Coefficients in bold font are significant with probability of 0.01, in italics with probability of 0.10, unemphasized coefficients are not significant. In model c), the path coefficient of *Peak Flow* on *Loc Choice* was fixed to the value 0.172.

model (figure 3.5a), except that the direct effect of population growth on urban growth is smaller. It is clear that while the constraining influence of hydrology on urban growth exists (as predicted by the theory), it *is* relatively modest: one should not rely on it solely to mitigate the effects of natural hazards on human settlements, as other factors (in particular accessibility) are sure to weigh more on location choice of urban agents.

A final relation of interest, the marginal effect of *Central* on *Urban*

Growth, should be highlighted. Within this path model (figure 3.5b), *Central* acts as a proxy variable for accessibility, which has a critical role in the microeconomic theory on urban location that has been adopted as an explanation for urban patterns. There were no firm expectations on the sign of this path coefficient, as both a positive and negative effect could be consistent with theory: a negative effect suggesting primarily greenfield expansion, i.e. the city becoming physically larger, and a positive effect suggesting urban growth through intensification (building of smaller dwellings in open spaces of the center or substitution of single housing units by taller buildings). Both dynamics could follow from an exogenous population increase. The resulting coefficient was positive but not statistically significant.

The urban patterns of Kampala, shown in figure 3.3, already suggest the existence of clear greenfield expansion in the borders of Kampala but also the intensification of built-up land cover in more central locations (particularly in 2016). It would appear the results of the second structural equations model (figure 3.5b, specifically the lack of significance of this path coefficient) is consistent with the presented assessment of change in the disaggregate spatial patterns: the greenfield development and the intensification of central locations may be canceling each other out, leading to a non-significant path coefficient of the arrow linking *Central* to *Urban Growth*. Since the urban growth data only reflects the horizontal expansion of built-up land cover, and not densification by multi-storey buildings, such effects cannot be further disentangled. Alternatively, it is also possible that an aggregate scale is too coarse to fully reflect the role of accessibility (through the proxy variable *Central*) in the urban growth patterns.

The latent variable model, figure 3.5c, results in a very poor fit to the data ($\chi^2 = 20.902$, $df = 7$, $p = 0.004$ [0.030]), up to the point of being rejected according to the χ^2 statistic. The path coefficients of the resulting structural equations model (figure 3.5c and equation 3.4) do not conform especially well to the data: while the path coefficients of *Pop Growth* on *Urban Area* and *Urban Growth*, as well as those of *Rain* and *Urban Area* on *Peak Flow* and of *Central* on *Urban Area* are all statistically significant and broadly in line with previous models (see figure 3.5), the path coefficient of *Central* on the latent variable *Loc Choice* is not significant, nor is the path coefficient of *Loc Choice* on *Urban Growth* (and the sign of the latter contradicts theoretical expectations, as it should be positively correlated to urban growth yet the estimated path coefficient is negative).

$$
\begin{array}{lclcl}
\mathit{UrbanArea} & = & 0.834 \cdot \mathit{PopGrowth} & + & 0.865 \cdot \mathit{Central} \\
 & & (< 0.001) & & (< 0.001) \\
\mathit{PeakFlow} & = & 0.209 \cdot \mathit{UrbanArea} & + & 1.980 \cdot \mathit{Rain} \\
 & & (0.001) & & (< 0.001) \\
\mathit{LocChoice} & = & -0.172 \cdot \mathit{PeakFlow} & - & 0.214 \cdot \mathit{Central} \\
 & & & & (0.341) \\
\mathit{UrbanGrowth} & = & 0.594 \cdot \mathit{PopGrowth} & - & 0.269 \cdot \mathit{LocChoice} \\
 & & (< 0.001) & & (0.463)
\end{array}
\tag{3.4}
$$

Synthesizing, the third structural equations model (based on a latent variable) does not seem to be a convincing causal explanation of the causal path models and should be rejected. Other path models are consistent with the data, as judged by the results of the structural equations model, and they verify the hypothesized relation between hydrological impacts (*Peak Flow*) and urban growth, a result which is also consistent with theoretical models of urban location choice and with previous case studies from the literature. They also characterize this relationship as statistically significant but relatively weak.

3.5 Discussion and conclusions

The causal path models developed (figure 3.2) reproduce, at an aggregate level, the process of urban growth caused by an exogenous population growth. They are meant to focus on the constraint posed by hydrological outcomes on urban growth. Statistically significant but relatively modest (negative) marginal effects of hydrological outcomes on urban growth were found, certainly smaller than the direct positive effects of population growth on urban growth (see figure 3.5).

From a methodological perspective, the main contribution of structural equations modeling is to capture pattern of covariation between all variables; particularly, one may derive theoretically important causal indirect effects, even if they are relatively weak (such is the case, for example, of the indirect effect of population growth on urban growth via urban area and peak flow discussed in section 3.4). Furthermore, the analysis of aggregate data presented is necessary before undertaking more detailed models (in terms of scale and of individuals vs. areas), since it provides a benchmark against which to test the emerging properties of complex modeling approaches. Indeed, if the spatial effects that are to be modeled from the bottom up do not manifest themselves at the top level, it would be very difficult to test such bottom-up model; this implies, though, that aggregate level explorations are necessary before delving into the impact on the system of individual traits.

This is not to say any level of aggregation can, or should, incorporate all key variables of a system. Accessibility to central locations was treated, in the structural equations models, by a categorical variable differentiating the more accessible hydrological units from the least;

yet it is a very rough measure of accessibility (for example, within each hydrological units, the variation of accessibility is very large). This approach was sufficient to noticeably improve the goodness of fit of the causal path (the model of figure 3.5b is a substantially better fit to the data than the model of figure 3.5a, which does not include the variable *Central*), but it did not provide sufficient information to replicate the location choice process through a latent variable (*Loc Choice* in the model reported in 3.5c). This is likely due to the high level of spatial aggregation of the data, necessary to focus on the impacts of hydrological systems but which obscures other basic relations such as the effect of accessibility. The ultimate lesson, then, is each phenomenon imposes on the analysis a scale and not all relations are evident at all scales. In this specific case, should the objective be to deepen the understanding of location choice, the methodology would have to incorporate at least two scales–a less detailed to reflect the impact of physical phenomena and a more detailed to capture the variation of accessibility.

A final methodological point regarding the assumptions of the causal path analysis is important and has thus far received little attention. The neoclassical microeconomic models which were used to construct the causal paths make critical assumptions to simplify the mathematical analysis that justifies much of their proofs. Among these, crucially, urban agents are assumed to be homogeneous: neoclassical microeconomic models cannot simultaneously account for variations of space (the environment in which urban agents make lcation decisions) and variation among the urban agents. This means, for example, that all urban agents are assumed to share a perception of risk (Filatova et al., 2011) or, even more importantly, all have the same income (e.g. see Glaeser, 2008, who incorporated two income groups into the basic model to explore public vs. private transportation but at the expense of simplifying other elements). Microeconomic models are not tractable in the absence of these simplifications. However, in the context of a relatively poor city with few institutional constraints, such as Kampala, this is likely a key methodological weakness, as the urban poor have strong incentives to appropriate land exposed to hazards that other urban agents would avoid. It may be an important factor contributing to the weakness of the relation between *Peak Flow* and *Urban Growth* in the results presented. Agent-based modeling has been proposed as a solution to these issues (Filatova et al., 2011); but, as commented, it is difficult to apply very stylized simulations in the absence of validation data for the relations these models must assume.

From a policy perspective, it is important to first acknowledge that the land use planning system of Kampala is particularly weak and ineffective, even in the context of Sub-Saharan Africa (Goodfellow, 2013 concluded the political bargaining environment of Uganda incentivizes political elites to render regulations impotent). It is tempting, then, to argue from the theory that flood mitigation should be a social consequence of land market equilibrium. While there is indeed an effect, it is relatively small in the broader context of important population growth. Therefore, land market dynamics cannot substitute a land use regulation system's functions in mitigating flood risk. This result is consistent with theoretical arguments put forward by Tatano et al. (2004), who argue land use regulations are necessary because of perception bias in vulnerability to natural disasters.

In the long run, though, the constraint of hydrological outcomes on urban growth can reinforce regulations. One can speculate that building constraints on flooded areas, which also coincide to a large extent with valley floors that would naturally be permanently flooded wetlands (i.e. areas that are for multiple reasons unsuitable for urban development), likely will generate less resistance among land owners and other urban agents. Such restrictions could be, thus, relatively easy to implement (although one must bear in mind Kampala presents a very complicated political and social dynamic, clearly intersected by land issues through the Kingdom of Buganda and *mailo* customary land tenure; this dynamic has a long history and its effects on modern Kampala, via land markets and urban patterns, has never been wholly understood). However, and perhaps more importantly, the existing constraints posed by hydrological outcomes on urban growth already delay urban growth in these unsuitable areas; in this sense, they contribute by allowing the city's administration time to set up and improve existing land use planning systems (a not trivial problem, given the deep social and political roots of the land use planning system's weakness, Goodfellow, 2013). Eventually, in the face of pressure from population growth, this constraining effect should fade. Has Kampala reached such a point already? Judging from the results presented, not yet; but one must caution that sustained population increase is changing urban growth patterns in Kampala: this can be seen in figure 3.3, for example, as the study area has already intensified substantially, to the point future development should mainly be greenfield beyond its limits. Additional effects of this population increase (such as non-simple interactions between urban growth, housing size, flood risk, and accessibility) should be expected to become increasingly important.

3.6 Appendix

Table 3A.1 Sample data used to estimate the structural equations models

ID	Catchment	Year	Pop Growth	Urban Area (ha)	Rain (mm)	Peak Flow (m^3/s)	Urban Growth (ha)	Central
1	Nakivubo	2016	11149	1792.6	89.1	91.147	92.8	1
2	Kansanga	2016	12016	865.3	89.1	30.467	78.2	1
3	Mayanja/Kalid.	2016	56455	1984.9	89.1	28.39	437.2	0
4	Kinawataka	2016	18431	1313.4	89.1	49.478	206.6	1
5	Nalubaga	2016	19442	757.3	89.1	31.6	242.4	0
6	Walufumbe	2016	12613	371.5	89.1	11.702	108.3	0
7	Lubigi	2016	69316	3233.3	89.1	93.046	430.7	1
8	Nalukolongo	2016	16167	1384.2	89.1	40.377	81.9	1
9	Nakivubo	2010	18135	1089.9	165.7	268.5	249.8	1
10	Kansanga	2010	11416	422.6	165.7	105.45	108	1
11	Mayanja/Kalid.	2010	41736	1204.2	165.7	83.23	410.2	0
12	Kinawataka	2010	12411	599	165.7	106.65	173.1	1
13	Nalubaga	2010	8954	330	165.7	117.83	128.6	0
14	Walufumbe	2010	6939	143.8	165.7	59.88	71.4	0
15	Lubigi	2010	51240	1721.7	165.7	310.01	403.9	1
16	Nalukolongo	2010	17690	574.3	165.7	135	145.4	1
17	Nakivubo	2005	8944	840	78.3	65.534	840	1
18	Kansanga	2005	6514	314.6	78.3	19.622	314.6	1
19	Mayanja/Kalid.	2005	23307	794	78.3	19.49	794	0
20	Kinawataka	2005	6522	425.9	78.3	35.276	425.9	1
21	Nalubaga	2005	4584	201.4	78.3	19.125	201.4	0
22	Walufumbe	2005	2178	72.4	78.3	6.53	72.4	0
23	Lubigi	2005	45306	1317.7	78.3	69.64	1317.7	1
24	Nalukolongo	2005	13014	428.9	78.3	30.971	428.9	1
25	Nakivubo	2016	10036	722.1	89.1	13.475	116.9	1
26	Kansanga	2016	16845	251.3	89.1	50.14	156.4	0
27	Mayanja/Kalid.	2016	22046	544.4	89.1	30.338	212	0
28	Kinawataka	2016	16962	334.6	89.1	51.142	224.7	0
29	Nalubaga	2016	7558	129.8	89.1	33.924	99.7	0
30	Walufumbe	2016	6223	50.1	89.1	18.239	54.1	0
31	Lubigi	2016	66475	881.4	89.1	10.145	528.8	0
32	Nalukolongo	2016	32132	305.3	89.1	58.51	242.9	0
33	Nakivubo	2010	8513	493.1	165.7	23.806	118.3	1
34	Kansanga	2010	5104	208	165.7	121.076	46.1	0
35	Mayanja/Kalid.	2010	5848	131.5	165.7	77.241	61.7	0
36	Kinawataka	2010	6079	283.1	165.7	176.475	86	0
37	Nalubaga	2010	3777	85.2	165.7	151.146	60.7	0
38	Walufumbe	2010	4201	65.3	165.7	37.38	45.7	0
39	Lubigi	2010	23954	552.1	165.7	36.846	212.1	0
40	Nalukolongo	2010	17986	485	165.7	167.656	159.2	0
41	Nakivubo	2005	4880	374.9	78.3	11.781	374.9	1
42	Kansanga	2005	3442	161.9	78.3	31.042	161.9	0
43	Mayanja/Kalid.	2005	2343	69.7	78.3	19.535	69.7	0
44	Kinawataka	2005	4675	197	78.3	35.222	197	0
45	Nalubaga	2005	673	24.4	78.3	19.347	24.4	0
46	Walufumbe	2005	461	19.6	78.3	9.409	19.6	0
47	Lubigi	2005	17001	339.9	78.3	9.859	339.9	0
48	Nalukolongo	2005	13094	325.8	78.3	42.792	325.8	0

Table 3A.2 Rainfall hyetographs of simulated events

Time		Intensity (mm/h) Year 2005 Total = 89.07	Year 2010 Total = 165.68	Year 2016 Total = 78.3
0	30	3.20	5.95	2.81
45	60	4.63	8.62	4.07
60	75	6.42	11.94	5.64
75	90	9.74	18.12	8.56
90	105	17.29	32.16	15.20
105	120	42.75	79.52	37.58
120	135	129.15	240.24	113.53
135	150	42.75	79.52	37.58
150	165	17.29	32.16	15.20
165	180	9.74	18.12	8.56
180	195	6.42	11.94	5.64
195	210	4.63	8.62	4.07
210	240	6.42	11.94	5.64
Daily total		89.1	165.7	78.3
Event total		75.9	141.2	66.7

4

A cellular automata model of urban growth to inform flood policy[1]

Abstract

Urban growth may intensify local flooding problems. Understanding the spatially explicit flood consequences of possible future land cover patterns contributes to inform policy for mitigating these impacts. A cellular automata model has been coupled with an *OpenLISEM* flood model to simulate scenarios of urban growth and their consequent flood; the urban growth model makes use of a continuous response variable (the built-up fraction) and a spatially explicit simulation of supply for urban development. The models were calibrated for Upper Lubigi (Kampala, Uganda), a sub-catchment that experienced rapid urban growth during 2004–2010; this data scarce environment was chosen in part to test the model's performance with data inputs that introduced important uncertainty. The cellular automata model was validated in Nalukolongo (Kampala, Uganda). The calibrated modeling ensemble was then used to simulate urban growth scenarios of Upper Lubigi for 2020. Two scenarios, trend conditions and a policy of strict protection of existing wetlands, were simulated. The results of simulated scenarios for Upper Lubigi show how a policy of only protecting wetlands is ineffective; further, a substantial increase of flood impacts, attributable to urban growth, should be expected by 2020. The coupled models are operational with regard to the simulation of dynamic feedbacks between flood and suitability for urban growth. The tool proved useful in generating meaningful scenarios of land cover change that incorporate policy for flood mitigation.

Keywords: Cellular automata, Flooding, Urban growth, Kampala, Model integration

[1] This chapter is based on: Pérez-Molina et al. (2017).

4.1 Introduction

The city of Kampala is growing rapidly, as exemplified by the accelerated increase of its urban footprint (Vermeiren et al., 2012). Because of a weak institutional setting (Goodfellow, 2013a) but also due to a complex physical context, this expansion has generated a number of negative impacts; among them, increased urban development has lead to greater runoff, a consequence of more impervious areas. As Kampala's drainage systems are inadequately developed and maintained – despite recent major investments in the system which have mitigated existing problems –, this, in turn, has contributed to aggravating local flooding (Lwasa, 2010).

The goal of this chapter is to present a geo-information technology-based tool to explore the flooding consequences of urban development. The selected instrument is a coupled urban growth-flood model. Emphasis is placed on the potential to create a diversity of meaningful land cover and flooding scenarios. These scenarios should respond both to policy and to social and physical factors which influence land cover patterns and their ensuing flood impacts. Further, the modeling approach must be tractable; given the complexity of emergent behavior in a city, this is achieved by adopting a simple, and therefore understandable, approach to geographic inputs. The urban growth model includes a strong component of randomness to account for seemingly irrational behavior by urban actors, such as poor enforcement of regulations or higher willingness, notably of the urban poor, to occupy and develop hazardous, flood-prone land.

4.2 Conceptual framework and previous work

The analysis approaches urban flooding in Kampala as a coupled human and natural system (Liu et al., 2007; Alberti et al., 2011). In the specific research context of flooding and landscape patterns of Kampala, urban growth is the main mediating phenomenon between human behavior and its physical consequences in terms of flooding. Figure 4.1 features both a breakdown of the elements within both subsystems and the hypothesized relationships between these.

Urban development is conceived as driven by an external demand for land (due to population growth); the location of urban development materializes based on three core principles: suitability of land, continuation of historical development, and neighborhood interactions (van Schrojenstein Lantman et al., 2011).

Suitability is determined by accessibility to urban centralities, following the approach originally developed by Alonso (Glaeser, 2008; Brueckner, 1987), which implies urban agents will choose to live as close as possible to these central locations. Further, physical characteristics of the land are also included as potential determinants – specifically, flooded areas (wetlands, permanently flooded, or flood zones, flooded as

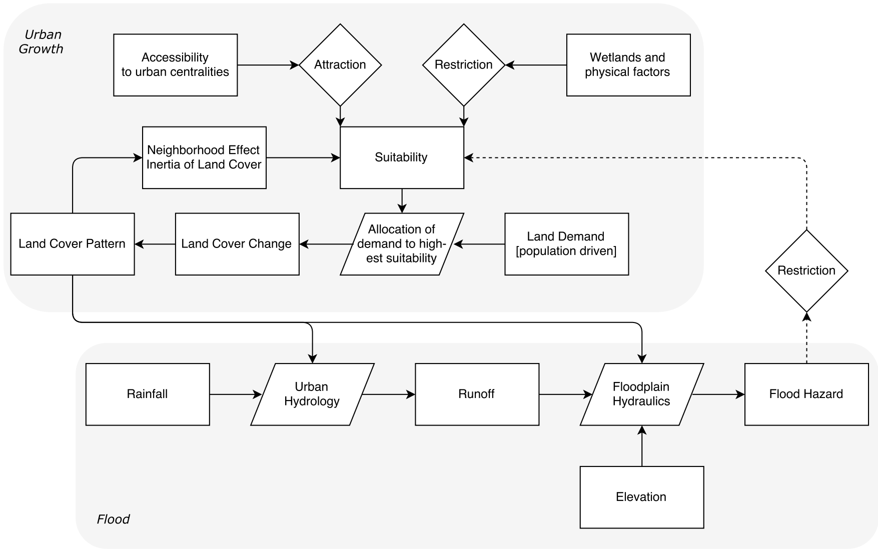


Figure 4.1 Conceptual framework: interaction between land cover and flood dynamics

a consequence of an extreme event) –, assumed to negatively impact the prospect of urban development (e.g., see Bathrellos et al., 2017). These physical factors follow from McHarg's original concept of suitability as an intrinsic and spatially specific characteristics of the land (McHarg, 1969).

The neighborhood effect assumes "transition[s] from one use of land to another is dependent on the land use of its surrounding cells" (van Schrojenstein Lantman et al., 2011, p. 38), a model originally proposed by Tobler (1979).

Flood, on the other hand, is thought of as a rainfall-runoff-flooding process: rain falls over the landscape, it is partly infiltrated into the soil an partly flows over the land until it reaches the drainage channels, where it may accumulate and, if so, results in inundation (Sene, 2010; Smith and Ward, 1998). This added flooding could change the suitability patterns; in particular, recurrently flooded locations should be less desirable for development (Bathrellos et al., 2017). While apparently obvious, this conceptualization assumes a short-run (event-based) view of flooding, a physical response with potential long-run consequences. However, the importance of this process – relative to other spatial factors determining suitability for urban land – is uncertain and may depend on the specifics of different sites.

Previous studies of the city of Kampala (Vermeiren et al., 2012; Abebe, 2013; Fura, 2013; Mohnda, 2013) have already quantified the impact of the most important determinants of its urban morphology, at mid and detailed scales. These approaches were all statistical, making use of

Logit econometrics, without spatial autocorrelation but including neighborhood effects as proxy variables of it (this methodological approach corresponds to the classical spatial statistical models used to investigate land cover and land use change; see Chomitz and Gray, 1994 for a full discussion on the conceptual and methodological decisions in this type of model). All show broadly consistent results: strong influence of the neighborhood in explaining transformation into built-up land cover, rapid growth rates of the urban footprint. The *OpenLISEM* integrated flood model, coupling surface hydrology (e.g., see Baartmans et al., 2012, Hessel et al., 2003, and Sánchez-Moreno et al., 2014) to a 2D flood model (Delestre et al., 2014), was used by Mohnda (2013) to assess various runoff reduction strategies in the Lubigi catchment, north of the Kampala central business district (CBD). Habonimana (2014) examined the flood model's sensitivity to input parameters and to spatially explicit representations of rainfall. More generally, the results of spatial-statistical urban growth models developed by Fura (2013) were used as inputs for the *OpenLISEM* flood modeling tool to estimate a diversity of scenarios, which included both future plausible land cover as well as interventions on the drainage system and alternative infiltration actions (Sliuzas et al., 2013).

The modeling exercise seeks to understand how diverse urban development strategies, especially land policy options, can lead to different flood dynamics. Upper Lubigi and Nalukolongo were selected as case studies, within the city of Kampala, Uganda, to develop the urban growth model and to test the integrated modeling ensemble (urban growth and flood models). The urban growth model is based on cellular automata, enhanced by including additional factors such as accessibility or physical constraints. An exogenoulsy determined land demand is allocated assuming that higher suitability locations are chosen by urban agents before lower suitability locations. The model was calibrated for the Upper Lubigi sub-catchment, 2004-2010, and validated using data from the Nalukolongo sub-catchment. Prospective urban growth scenarios were then defined and simulated for Upper Luibigi, 2020; the runoff and flooding patterns of these scenarios were assessed by applying the *OpenLISEM* flood modeling tool using the scenarios as inputs.

The developed modeling approach integrates urban growth and flood modeling into a single ensemble, supporting the possibility of exploring feedback effects between human and environmental systems. Further, the urban growth model has been designed with a continuous response variable (the percentage of built-up land cover), and it separates a spatially explicit simulation of potential supply from a suitability map controlling where is development more likely; this provides flexibility to replicate highly random trends (like urban growth in Kampala) as well as design-oriented supply scenarios. Since data of only two time periods is available, a second catchment – similar to the calibration location – was selected to validate the model. Finally, the case study has been selected partially to explore if remotely sensed and field measurement data can contribute to mitigate data scarcity and quality problems.

4.3 Methods: CA calibration, validation, and integration with the flood model

A cellular automata model was developed to simulate urban growth in the Upper Lubigi sub-catchment. In contrast to typical cellular automata models, it is based on continuous variables to characterize land cover cell state such that all relevant categories of land can be simulated concurrently; the use of such continuous-variable approach is more broadly discussed in van Vliet et al. (2012). The model was calibrated using land cover data from 2004 and 2010. Model development resulted from identifying a set of potential determinants of land cover change, from the theory on urban location (Alonso's model, McHarg's suitability concept, agglomeration dynamics, see Geographical Sciences Committee and others, 2014) and from previous statistical studies of urban growth in Lubigi (Abebe, 2013; Fura, 2013). Information to derive land cover data models was only available for two periods (2004 and 2010). The land cover change model was formulated and calibrated using only data from the Upper Lubigi sub-catchment. The lack of a third period precluded model validation on Upper Lubigi; thus, the second sub-catchment (Nalukolongo) was added to validate the calibrated urban growth model on exogenous data.

A set of eight factors was identified, each a potential determinant:

- Neighborhood factor: average built-up percentage within a moving window of 3×3 cells, estimated from the previous period's cumulative allocated development.
- Random factor: random number between 0.00 and 1.00, representing irrational behavior by urban agents (developers, home and land owners, etc.)
- Travel time to CBD: a reflection of regional accessibility and possible travel preferences by urban agents.
- Travel time to nearest subcenter: a reflection of regional accessibility and possible travel preferences by urban agents.
- Wetland factor: a representation of physical unsuitability for building in naturally flooded soils.
- Non-vegetation percentage: introducing inertia (land cover changes tends to be stable; specified as non-vegetation to avoid endogeneity problems caused by directly using built-up and on-road bare soil percentages); estimated from the previous period's simulated land cover.
- Road density: based on 2004 data, it is an index of local accessibility.
- Flood depth in base year (2004), accounting for unsuitability of flooded areas for urban development.

Auxiliary models, using each a single factor as determinant, were produced. All models incorporate a correction for institutional land uses that tend to change very little: if land use was institutional according to a 2002 land use map, no change is assumed to occur in those cells. From understanding the effect of each factor, separately, on the urban growth pattern, a combination of them was selected, guided by theories on location of urban activities and the predictive power of alternative modeling formulations. Different weight combinations for the spatial factors were tested, selecting the one best replicating the 2010 land cover map of Upper Lubigi.

The resulting calibrated model (i.e., set of spatial factors and their weights) was used to predict urban growth patterns in the Nalukolongo sub-catchment, since – as noted – no additional land cover data of the Upper Lubigi area was available for validation.

The model was coupled with an flood model (a calibration of the *OpenLISEM* flood modeling tool) of Upper Lubigi, calibrated by simulating a $100mm$ rainfall event. In the absence of discharge rate data for primary and secondary drainage channels, calibration of this flood model was based on: field observations, interviews with local residents on past flood impacts (from Oct. 2012), and comparison with model outputs of the Kampala Drainage Master Plan (Sliuzas et al., 2013).

4.3.1 Geographical setting

The study area comprises two sub-catchments located within Kampala (see figure 4.2). They have a combined area of $44.2km^2$ (63.4% is occupied by Upper Lubigi and 36.6% by Nalukolongo). Both correspond to the upper reaches of a hydrological system that drains inland, as opposed to directly into Lake Victoria. They also include peripheral but important subcenters of the city. Upper Lubigi, in particular, experienced very rapid growth during the 2004-2010 period: the Northern Bypass, a major regional road, was completed in 2005, attracting substantial development in the following years.

Previous research efforts on modeling urban growth and its impact on flooding (Sliuzas et al., 2013), have concentrated on the Upper Lubigi sub-catchment; current results reported in this paper have built on these efforts, developing more tightly coupled models of Upper Lubigi.

4.3.2 Flood modeling: calibration of *OpenLISEM* for Upper Lubigi

Flood modeling was implemented using the openLISEM flood modeling tool (Jetten, 2018). Originally developed as an erosion model (De Roo et al., 1996), it has been recently extended to simulate event-based floods. It is a rainfall-runoff-flood model, which replicates physical processes at very detailed temporal resolution (0.1-60 seconds) for mid spatial resolution catchments (between 5 and 50 m cells), with the aim of assessing flood hazard for decision support. *OpenLISEM* has been tested in urban catchments of Sub-Saharan Africa (in addition to Kampala, efforts are underway to model selected catchments in Kigali, Rwanda, e.g. Habonimana, 2015) and on four Caribbean islands to analyze flood

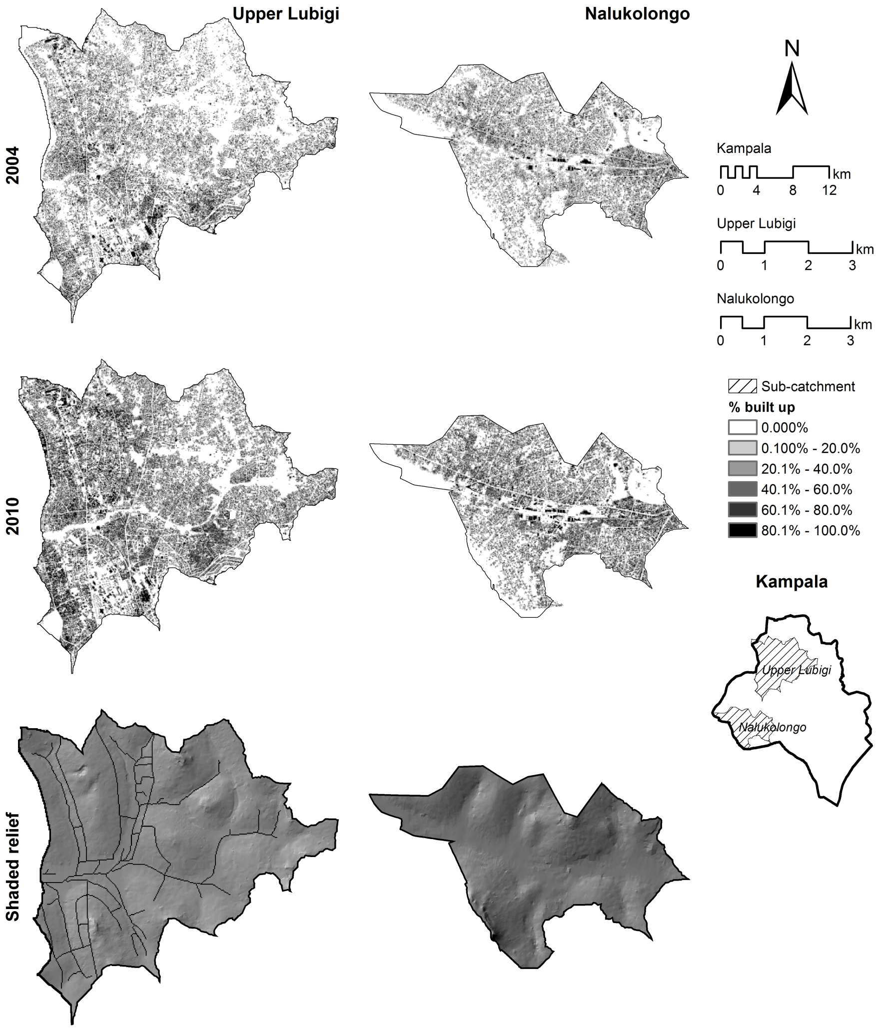


Figure 4.2 Study area: 2004 built-up percentage in Upper Lubigi and Nalukolongo sub-catchments in Kampala, Uganda

hazard, as part of the CHARIM project (van Westen et al., 2015; Jetten, 2016).

OpenLISEM simulates the consequences of an extreme rainfall event in terms of runoff and flood. In the first phase, runoff is determined and accumulated towards river channels (i.e., downhill) using a kinematic wave approach over a predefined network (this process corresponds to urban hydrology in figure 4.1). At a second phase, water is routed through the channels with a 1D kinematic wave[2]; channel overflow resulting in flooding is modeled as overflow from the river channels towards the higher elevations of the floodplain, using a 2D approximation of the shallow water equations (Delestre et al., 2014). The kinematic wave functions for overland flow and for the channels are solved using a four-point finite-difference solution.

The characteristics of the terrain are incorporated into *OpenLISEM* as cell values in a series of input raster maps. These maps identify the location, width and slope of channels, characterize the infiltration properties of soils, describe interception and imperviousness (based on land cover), and provide elevation data, used for flow routing. The simulated physical processes for the Upper Lubigi subcatchment were:

- Rainfall. The rainstorm event is represented by a map of discrete units, each of which has an associated hyetograph. For the Upper Lubigi sub-catchment, a $100mm$ magnitude event –approximately equal to a $1:10$ year return period – with measurements every $10min$ was selected (Sliuzas et al., 2013). This event was applied to the entire sub-catchment.
- Infiltration. The fraction of rainfall infiltrated into the soil is simulated using a 1 layer Green and Ampt approach (Kutilek and Nielsen, 1994), which uses saturated hydraulic conductivity ($Ksat$, in mm/h), porosity, and an initial moisture content of the top soil layer. For each time step, the rainfall intensity is compared to the infiltration rate to produce runoff. The infiltration characteristics of Upper Lubigi were sampled; the results are synthesized in table 4.1.
- Land effects. Elements on the surface of land (on top of the soil) interfere with infiltration in two ways, by storing a fraction or by preventing infiltration.
  - Canopy storage. A fraction of rainfall that does not reach the soil because it is stored in the vegetation; it is a function of the type of vegetation and the fraction of vegetation land cover in every cell within the study area. Interception was estimated as: $S = 0.59 \cdot LAI^{0.88}$, with LAI (the leaf area index) calculated from: $LAI = -\ln(1 - VegFraction)/(0.4 \cdot VegFraction)$ (De Jong and Jetten, 2007).
  - Impervious land cover. Certain areas covered by specific land cover classes (buildings, tarmac) are identified as the percentage of each category within every cell. The area of each cell corresponding to these land cover categories is assumed to have 0 infiltration rate; this water volume is, therefore, added to the non-impervious part of the cell.
- Overland flow and channel flow. The kinematic wave uses the flow velocity based on the Manning formula. Manning's n values for flow resistance were estimated from a baseline resistance increased by

[2] Alternatively, and as implemented in other chapters of this dissertation, *OpenLISEM* can also route water using a 2D kinematic wave or a 2D diffusion process.

Table 4.1 Soil properties

Unit	Porosity (cm^3/cm^3)	Bulk Density (kg/m^3)	$Ksat$ 1/ (mm/h)	n
Vegetation (grass, shrub)				
Valley floor (C)	0.58	1210	1.9	5
Bottom slope (SCL-SC)	0.52	1410	16.8	9
Mid-slope (SCL)	0.54	1460	2.8	2
Hill top (SCL-SL)	0.52	1410	37.6	5
Bare soil (compacted)				
Valley floor (C)	0.57	1310	0.0	8
Bottom slope (SCL-SC)	0.55	1490	6.5	5
Mid-slope (SCL)	0.58	1550	3.3	2
Hill top (SCL-SL)	0.54	1470	2.9	4

C: clay, SCL: sandy loam clay, SL: sandy loam
1/ Saturated hydraulic conductivity

the effect of the grass cover (C): $n = 0.05 + 0.1 \cdot C$. Furthermore, the effect of buildings (built-up land cover fraction, Cb) increases resistance by: $n = n + 0.5 \cdot Cb$. The n value for the main channel was set to $n_{chan} = 0.025$ for a smooth, straight, bare soil channel. The effect of obstructions and garbage in the channels could not be taken into account.

Land cover data determines impervious land cover, canopy storage, and resistance to water flow. The landscape should be described in terms of categories corresponding to these processes: impervious land cover is relevant for built-up land cover, tarmac, and unpaved roads – which, because they are compacted soil, have lower infiltration rates –; for canopy storage, the amount of vegetation is the key determinant; off-road bare soil is also important because it indicates areas where the infiltration rate is applied, as opposed, for example, to open water, where no infiltration occurs. Thus, five categories have been chosen to describe the urban landscape of the study area: built-up, tarmac, unpaved roads, off-road bare soil, vegetation, and water.

4.3.3 Urban growth modeling of Upper Lubigi

An urban growth model, based on cellular automata, was developed to describe land cover change processes in Upper Lubigi. Space within the sub-catchment was idealized as an array of square cells, of $20m$ side, each cell being an automaton A characterized by a set of states (**G**), a set of transition rules (**T**) governing changes to these states, and a set of states of neighboring cells (**R**):

$$A \sim (\mathbf{G}, \mathbf{T}, \mathbf{R}) \tag{4.1}$$

The state **G** of the automaton A is defined as a set of continuous variables encoding the percentages of land cover categories selected in subsection 4.3.2. Changes to this state are triggered by urban development, which is an increase in the built-up fraction

The transition rules **T** define the state of the automaton ($\mathbf{G_{t+1}}$) in period $t+1$, based on the automaton's state ($\mathbf{G_t}$) in the preceding time step, and on an input, I_t (also corresponding to the preceding time step):

$$\mathbf{T} : (\mathbf{G_t}, I_t) \rightarrow \mathbf{G_{t+1}} \tag{4.2}$$

This input, I_t, is a measure of the development potential for built-up land cover, taken to be a representation of urban land uses' location.

I_t is defined as the weighed summation of several factors, each factor normalized between 0 and 1: (1) the neighborhood factor, defined for each cell as the average percentage of built-up land cover within a Moore neighborhood (i.e., the 8 cells surrounding it), (2) accessibility to urban centralities (estimated travel time through the network to the CBD), (3) physical factors restricting urban development, specifically an index assigning a value of 0 to permanent wetlands, 0.5 to seasonal wetlands, and 1 to non-wetland cells, and (4) non-vegetation percentage, a proxy for existing built-up in the previous period (but which, unlike the built-up percentage, mitigates potential endogeneity problems).

Following Yeh and Li (2001) and Li and Yeh (2000), the amount of built-up land cover change is assumed to be additive – because in developing countries, cities generally expand by greenfield projects, which shows up in land cover maps as an additional fraction of built-up land cover.

$$G\,(1)_{i,t+1} = G\,(1)_{i,t} + \Delta G\,(1)_{i,t} \tag{4.3}$$

where $\Delta G(1)_{i,t}$ is the increase in built-up percentage. Therefore, for each category, the model assigned:

$$\Delta G\,(1)_{i,t} = \begin{cases} SimDem_i, & \text{if } I_t \in [k, n] \\ 0, & \text{if } I_t \in [0, k] \end{cases} \tag{4.4}$$

such that:

$$\sum_{i=k}^{n} \Delta G\,(1)_{i,t} \approx LD_{t\ to\ t+1} \tag{4.5}$$

with $SimDem_i$ the simulated demand alloted to cell i if this cell's state changes, n the total number of cells in the study area, and $LD_{t\ to\ t+1}$ the total land demand for that land use/land cover category in the time interval t to $t+1$.

Since observed development clearly expands over space, for each simulation run, a rule was introduced: that a cell can only change once during the entire simulation, even if it is more suitable than others. This ensures the same cell will not be developed until saturation. The calibration period was six years; therefore, prospective simulations are performed for five period intervals (e.g., to simulate urban growth for 2010-2020, two sequential simulations are completed: 2010-2015 and

2015-2020; the results of the 2010-2015 simulation are used as inputs for the 2015-2020 simulation).

The amount of change, $SimDem_i$, of each cell is determined by examining land cover change maps of the sub-catchment; it is a function of the total land cover change occurring in a 6 year period, randomly distributed in space: built-up land cover change of 2004-2010 was averaged across cells and multiplied times a spatially random map (with values between 0.00 and 1.00) and times an expansion factor equal to 3.50; this expansion factor is introduced to control for the intensity of development. Since the average of a random uniform map (varying from 0.00 to 1.00) is 0.50, an expansion factor of 2.00 implies the intensity of development follows a purely random pattern; expansion factors greater than 2.00 should be interpreted as increasing the level of agglomeration in new development. $G(1)_{i,t=baseline} + SimDem_i$ is assumed to have a maximum value of 0.85.

Other land cover categories are updated based on the change in built-up land cover. The increase in unpaved road fraction is assumed to be proportional to the ratio of unpaved road to urban growth of the baseline year of the cell. Should the sum of unpaved road and built-up land covers exceed 1.00, unpaved road is taken as equal to $1 - frBuiltUp$ and the cell is assumed to be totally urbanized. Water and tarmac roads are defined as external to the simulation, meaning they are initial conditions that do not vary in time. When these land covers are present in a cell, allocation is calculated for the non-paved and non-water percentage. Vegetation and off-road bare soil are allocated to the fraction remaining after updating all other land cover categories. Such fraction is distributed proportionally to baseline year ratio of vegetation to off-road bare soil.

Constraints on institutional land uses (as identified in the 2002 land use map of Kampala), which are not expected to suffer changes, are introduced by setting $SimDem_i = 0$ for the areas occupied by these land uses.

4.3.4 Scenario descriptions of Upper Lubigi

Two scenarios were specified for the Upper Lubigi sub-catchment to assess a land use policy intervention, stringent protection of wetlands. These scenarios were projected for a 10 year period (adopting as the baseline year 2010).

Population projections for the city of Kampala were derived from UN statistics (United Nations, 2018); the fraction of population growth for Upper Lubigi was estimated by multiplying the percentage of population living in the Kawempe division (22% in 2002), based on census data. Land demand was estimated by dividing population growth into gross population density of the baseline year (2010). Total land demand for the 10 year period was estimated at $426ha$ (Pérez-Molina, 2014).

Spatial allocation of land demand was based on the calibrated urban growth model (i.e., the choice of factors and weights, product of the calibration procedure). The trend scenario (P01) was calculated with unconstrained supply of development ($SimDem_i > 0.00$) for wetland areas.

A defensive policy (P02), no future urban growth in wetland areas, was simulated by setting the amount of change – $SimDem_i$ – to 0.00 for cells within them.

Both scenarios assume improvements on the main drainage channel, completed in 2013, were operational in 2020. Neither takes into account potential maintenance problems, especially silting, which progressively reduce the capacity of these drainage channels in Kampala.

4.3.5 Model integration

Model integration describes the transmission of information between modeling components. Computationally, model integration can range from the so-called "loose coupling" (in which data is exchanged between models using relatively simple data formats) to "tight coupling", which occurs when the capabilities of one model are included into another (McColl and Aggett, 2007). An extended discussion of model integration can be found in Sui and Maggio (1999).

The models presented in this paper, while not fully (tightly) coupled, are an advance on loose coupling. *OpenLISEM*, the selected rainfall-runoff model, was developed using the PCRaster software (Karssenberg et al., 2010) for dynamic and spatially explicit environmental modeling (combined with C++). Indeed, data preparation for an openLISEM run is carried out using a PCRaster script (Jetten, 2014). Accordingly, all digital maps use the .map file format of PCRaster maps.

The urban growth model described in section 4.3.3 and the scenarios of section 4.3.4 are all the result of combining data using the PCRaster platform. Consequently, there is no need for data conversion for the loose coupling between the urban growth and flood models of Upper Lubigi. The basis of coupling consists in renaming the outputs of the urban growth model in such a way that the flood model can recognize them.

The application presented follows the pseudo-code summarized in algorithm 1. The urban growth model was used to simulate the future land cover of Upper Lubigi. This spatially explicit land cover forecast of 2020 was then evaluated with the *OpenLISEM* flood modeling tool.

4.4 Results and discussion: calibration, validation, and prospective simulation

4.4.1 Cellular automata model calibration

The cellular automata model calibration was achieved through an incremental approach to the introduction of information. In a first phase, a set of auxiliary models were run, each using a single spatial factor as suitability criterion. This yielded three dynamic models (based on the neighborhood, non-vegetation percentage, and random factors) and five static models. The results were assessed to understand the effect

Algorithm 1 Model execution with no feedback between urban growth and flood

Execute PCR UGM t:[2010,2015] {UGM designates the cellular automata urban growth model. t:[2010,2015] indicates the script runs from 2010 to 2015. Input: 2010 land cover maps}
Execute PCR UGM t:[2015,2020] {Input: 2015 simulated land cover}
return Built-up, vegetation, tarmac, unpaved road maps for 2020
Execute PCR Data compilation script t:[2020] {Input: 2020 simulated land cover and physical maps of sub-catchment}
Execute openLISEM ULFlood t:[2020] {ULFlood is the calibrated flood model for Upper Lubigi. Input: compiled data}
return Max. flood depth and catchment flood results for 2020

of differing urban dynamics on urban growth patterns. Thus, for example, if allocation is based solely on the travel time to the CBD, then the growth pattern should concentrate on the southern areas of Upper Lubigi – those closest to the CBD (see figure 4.3). Some interesting insights can be gleamed from such analysis; for example, both the neighborhood factor and the wetlands factor seem to drive development away from the flood plains.

Factor selection was based on empirical and theoretical considerations. From observing 2004-2010 urban growth patterns, it appears there is a strong influence of inertia and local agglomeration: new development tends primarily to locate in the western-most area of Upper Lubigi, where current urban land uses are denser. Spatial statistical models of urban growth in Upper Lubigi and Kampala support this, by finding neighborhood factors are strong predictors of change in built-up land cover (Abebe, 2013; Fura, 2013). Wetland areas also seem relatively free of development, although it is unclear whether this should be attributed to inertia/agglomeration effects (lack of existing development in the core wetland) or to physical suitability constraints. Travel time to the CBD was added as a theoretically central dynamic; at the Upper Lubigi scale, this would be manifested as a pattern of urban consolidation in the south (nearer the city center) preceding further development in the north of the sub-catchment.

Factor weights were selected relative to the neighborhood effect: travel time to CBD and non-vegetation factor were deemed less strong and wetlands, most strong (evidenced in the generally clear wetland core areas in aerial imagery, although encroachment does occur in the fringes). Non-vegetation is, to a point, redundant to the neighborhood effect, since both factors tend to favor locations with existing development.

Accessibility as a determinant, on the other hand, is weak probably due to widespread congestion in Kampala, which reduces the advantages of main roads over local unpaved roads (accessibility was estimated under near free flow conditions). This likely weakens any accessibility effects. The final selected weights were: 1.00 for the neighborhood factor,

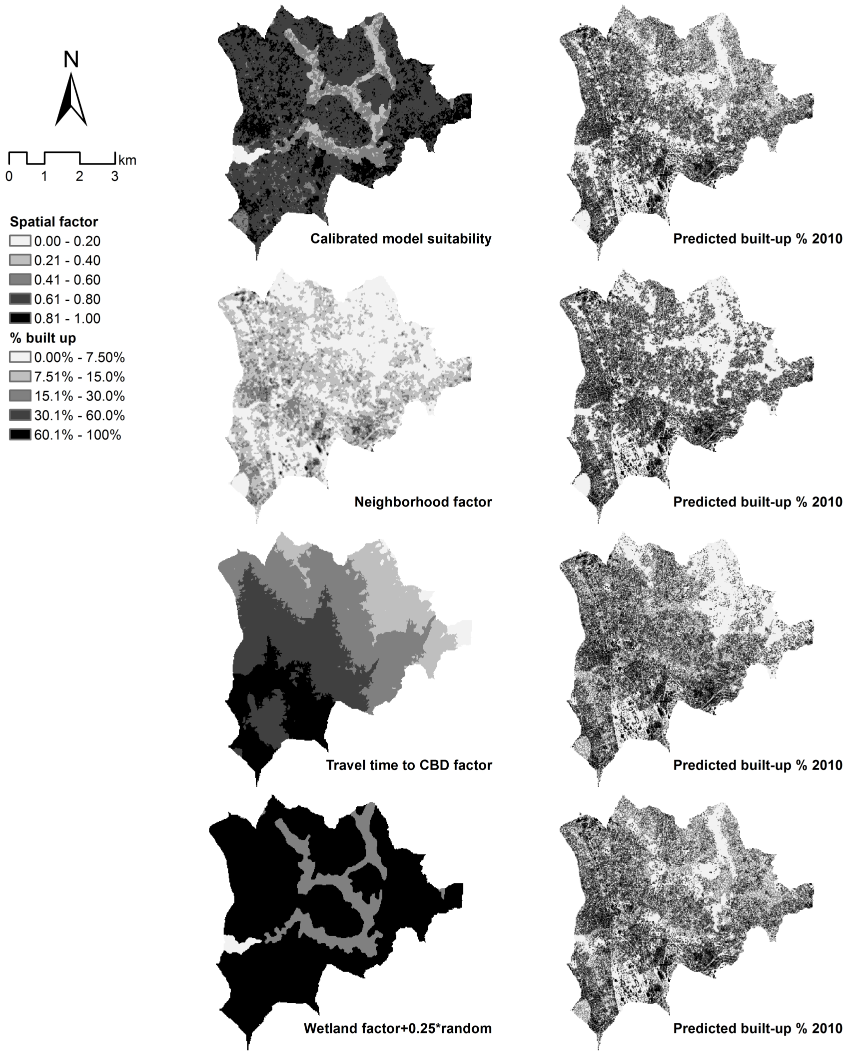


Figure 4.3 Calibration: selected spatial factors and auxiliary model results

0.500 for non-vegetation and travel time to CBD factors, and 2.00 for the wetland factor.

As can be seen from figure 4.4, the composite suitability (S01) performs better than simpler versions (S02, an equal weights average of the same four factors, and S03 to S10, the single factor models). There is hardly any difference in per pixel terms, but this is not surprising, as urban growth models are better at pattern prediction than individual location – and in fact it is not very useful to predict exact locations; general patterns are much more relevant (van Vliet et al., 2013).

Looking at other measures, some single model factors perform slightly

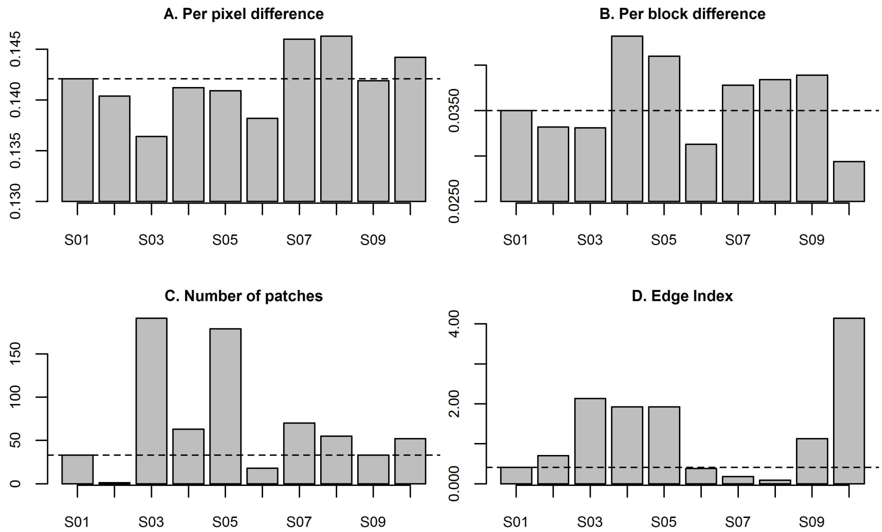


Figure 4.4 Calibration: assessment of auxiliary cellular automata models. Difference between calibration scenario and 2010 land cover map measure [For 2010 land cover map, number of patches: 1736; edge index: 64.10]

S00: 2010 land cover map, S01: composite suitability-weighed summation, S02: four factor summation-equal weights, S03: neighborhood factor, S04: travel time to CBD, S05: travel time to subcenters, S06: non-vegetation factor, S07: wetlands $+0.250 \times random$, S08: flood depth $+0.250 \times random$, S09: road density, S10: random factor.

better than the composite (S01) in terms of per block difference (S03, neighborhood factor; S06 non-vegetation factor, and S10, random factor). This clearly suggests the importance of inertia, which is also partially measured by local agglomeration (i.e., the neighborhood factor) in determining urban growth of Upper Lubigi. The neighborhood factor (S03), however, is not a good predictor for number of patches nor does randomness (S10) properly simulate the edge index. One could argue the non-vegetation factor does outperform the composite index; however, a model based solely on this factor tells a much oversimplified story of urban growth, one in which only locations with existing development intensify. While this is part of the general picture, prospectively it oversimplifies by ignoring greenfield development: a dynamic that will become increasingly important as space is fully occupied in existing urban areas. It also misses out on wetland encroachment from informal development; this could be a major problem looking forward. Because of this, for prospective modeling, it is best to choose the more complex composite suitability index.

The selection of factors resulting in S01 permits the replication of the observed urban growth process of Upper Lubigi in morphological terms.

But it does not address the inherent randomness of urban growth in the study area. Figure 4.5 shows the comparison of the model specified for S01 to eight models (S11 to S18) in which, to the four factors selected for S01, were added a random term with weight increasing from 0.500 to 2.00 in 0.250 increments.

As can be seen in figure 4.5, there are no clear trends: randomness definitely does not improve the per pixel and zone differences with respect to the 2010 land cover map. For the edge index, with exception of S16 and S18, there seems to be a weak bell shape, with lower differences for the extremes (S11, randomness weight of 0.500, and S17, weight of 1.75).

When analyzing the number of patches, evidence is less systematic – perhaps because all differences are relatively small (the number of patches of the 2010 land cover map is 1740, so the difference with respect to it are two orders of magnitude lower). However, S17 is one of two scenarios (the other being S14, weight of randonmness equal to 1.00) for which the difference in the number of patches with respect to the land cover map of 2010 is lower than for the calibrated model. Accordingly, the randomness factor was incorporated with a weight of 1.75, based on its improvement of quantitative measures of prediction.

In terms of substantive interpretation, one can justify including randomness as a reflection of informality, poor land use regulation, uncertainty of assumed dynamic relationships (e.g. accessibility is strongly affected by congestion, a factor not incorporated into the model), etc. To ascertain how important it is, the quantitative assessment of the calibration scenarios should be expanded substantially – since the impact of introducing randomness is not intuitively evident and visual inspection of resulting maps is not a straightforward comparison method. But such an exercise is beyond the scope of this chapter.

4.4.2 Cellular automata model validation

Model validation results from analyzing the urban growth process for 2004-2010 of the Nalukolongo sub-catchment. In the absence of land cover data for a third period in Upper Lubigi, a second catchment within the city of Kampala was chosen and used as a test case. This sub-catchment shares characteristics with Upper Lubigi such as being an inner suburb, including substantial wetlands, and having a main road connecting it to the city center. It also presents some differences, such as greater consolidation of urban areas and less undeveloped land.

Built-up land cover expansion was simulated for 2010 using 2004 as a baseline year, and the factors and weights selected in the calibration section: neighborhood effect, non-vegetation percentage (weight equal to 0.500), travel time to CBD (weight equal to 0.500), wetlands factor (weight equal to 2.00), and randomness (weight equal to 1.75). The simulated built-up pattern is reported in figure 4.6.

As can be seen, the resulting pattern is remarkably similar to the 2010 land cover map. The main differences are the greater predicted

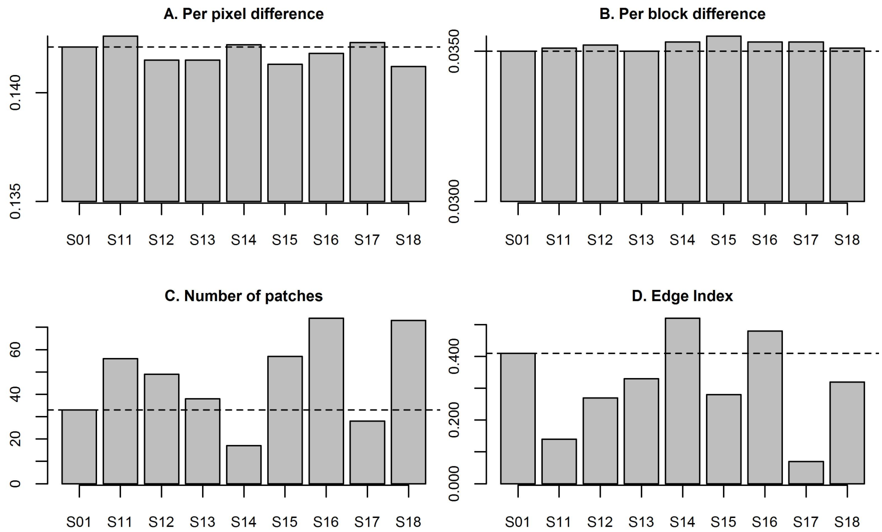


Figure 4.5 Calibration: introduction of randomness. Difference between calibration scenario and 2010 land cover map measure [For 2010 land cover map, number of patches: 1736; edge index: 64.10]

S01: composite suitability-weighed summation. Random factor weight increases from 0.250 to 2.00 by 0.250 increments from S11 to S18

growth in the southwest limit, relative to the land cover map, and the lower predicted growth in the northwest.

Further, the land cover map of 2010 shows a greater degree of organization, in the sense that streets are easier to detect, whereas the simulated map includes a scattered urban growth pattern. When analyzing the landscape metrics (figure 4.6), one can see the urban growth model is more successful in simulating Nalukolongo's development than for Upper Lubigi (except for patch complexity, measured through the edge index): per zone difference is 2%, as opposed to over 3% for Upper Lubigi; the difference in number of patches is 20 (out of 1031 for the land cover map) and the edge index difference is 0.10 (with a land cover map index of 49.1); contrast this to over 20 patches of difference (out of 1736) and less than 0.10 for the edge index (65.1 for the 2010 land cover map) for Upper Lubigi.

In synthesis, the calibration procedure was successful, as judged by applying the resulting model (factors and weights) to an independent data set (the Nalukolongo sub-catchment).

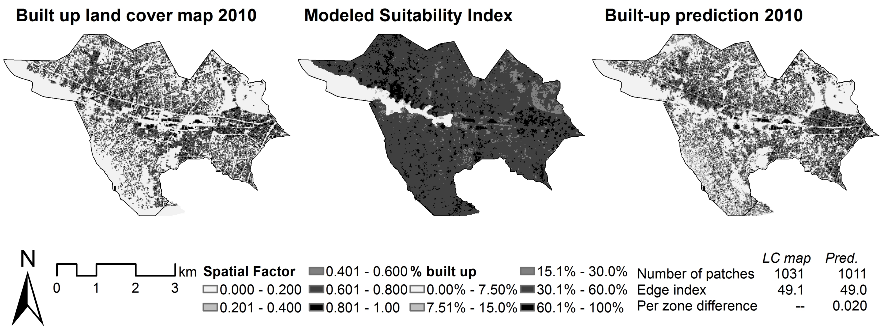


Figure 4.6 Validation: 2010 land cover map and simulated land cover. Landscape measures report the differences between validation scenario and 2010 land cover map measure

4.4.3 Prospective simulation of the Upper Lubigi sub-catchment

An exercise in prospective simulation has been undertaken to demonstrate the possibilities of the modeling approach. The spatial factors and weights of the calibration and validation results (a neighborhood factor with a weight of 1.00, travel time to CBD with weight of 0.50, wetland factor of weight of 2.00, non-vegetation percentage with weight of 0.50, and random factor with weight of 1.75) was used to compute the scenarios. A defensive landscape strategy, no further development allowed in wetland areas, was simulated for 2010-2020. The resulting hydrological impacts were compared to trend conditions.

The results of the prospective simulations should be cause for concern in the city of Kampala. In the face of a weak institutional position, urban managers at the Kampala City Council Authority have attempted to control urban encroachment into wetlands. Yet the simulation results strongly suggest this strategy is not effective: predicted total discharge volume, peak discharge flow, percent of rainfall discharged at the outlet, and flooded volume at maximum flood level are all nearly identical for scenario P01 (trend conditions) and P02 (wetland protection) at the year 2020 (see figure 4.7).

Even more problematic, relative to the baseline year conditions (scenario P00, 2010 land cover map), hydrological impacts are forecast to deteriorate substantially. Total discharge and discharge to rainfall ratio are predicted to increase 37%, flood volume at maximum flood level by 44%, and peak discharge flow by 4%. Perhaps the only positive finding is, flood patterns seem nearly identical for all three scenarios (figure 4.7), which would mean current conditions will worsen but the rise of new problematic spots will be limited.

An examination of flooded built-up area (table 4.2) also shows a significant impacts. Total flooded built-up area increases by 59% from baseline year conditions to 2020 scenarios –both of which predict nearly

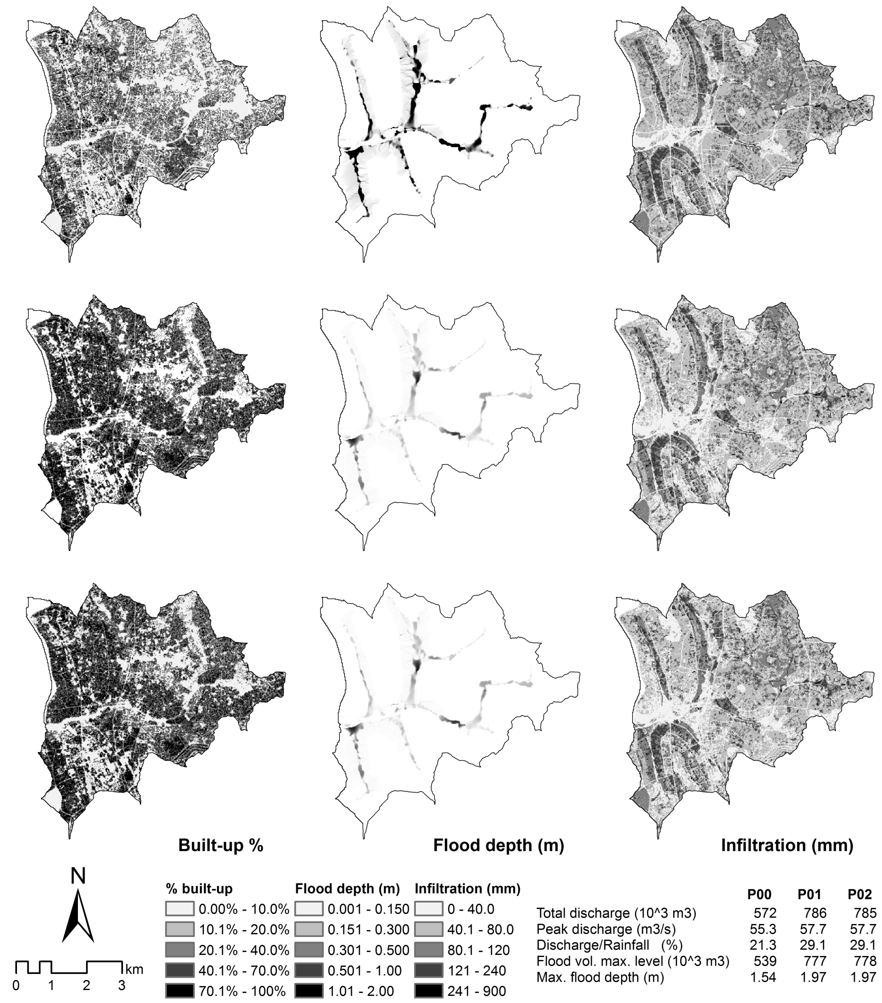


Figure 4.7 Simulation: prospective urban growth for selected scenarios

P00: 2010 land cover map, P01: 2020 trend, P02: 2020 wetland

identical flooded areas. More troublesome, the severely flooded built-up area (over 15*cm* of flood) almost doubles, from 22 to over 40*ha*. These results prove that the increase in flood impact should be attributed to greater flooding, rather than encroachment into flooded areas (since the policy scenario P02 assumes no development happens within wetlands).

In conclusion, given flood problems already exist in Upper Lubigi, more aggressive actions will certainly be required to mitigate the flood-related consequences of expected urban growth.

Table 4.2 Flood impacts on built-up land cover

	P00 Base year	P01 Trend	P02 Wetland
Flooded area total (ha)	534	588	588
Flooded area over 15cm (ha)	114	146	146
Infiltration (mm)	75.9	67.5	67.6
Built-up total flooded (ha)	139	222	222
Built-up flooded over 15cm (ha)	22.0	40.3	40.3

4.4.4 Conceptual and computational model integration

Conceptually, model integration depends on the phenomena themselves and whether they are well suited to each other in terms of scale and extent. In this sense, the compatibility between urban growth and flood events is a precondition to computationally integrate the models of these phenomena.

Elga et al. (2015) analyzed the problem of scale in hydrological modeling of urban catchments. They found that hydrological models developed to examine the relationship between land use change and hydrology concentrated on infiltration and runoff production. Time scales for runoff generation were found to range from minutes to days and spatial resolution, from 10^{-1} to 10^{3} m (Elga et al., 2015, p. 67). Depending on the complexity of soil patterns, infiltration could require very detailed information: a spatial resolution of up to 1 m, with additional vertical heterogeneity, especially in areas where soils have been disturbed by human agency.

Urban processes operate at coarser spatial scales and are, in consequence, discernible only at coarser time scales. In their review of urban attributes, Cowen and Jensen (1998) concluded USSGS Level I and II land use/land cover – level I aggregates all built-up into a single category, level II disaggregates into residential, commercial/services, industry, etc.; see Anderson (1976) – require a minimum temporal resolution of 5 to 10 years and spatial resolution of 5 to 100 m. While higher resolutions could generally be preferred, they also lead to greater object diversity (Bhatta, 2010), which causes problems for algorithm-based analysis.

The application described in subsection 4.4.4 presents no conceptual integration problems. Flood depth, in the case of Upper Lubigi, was not found to be a determinant of land cover change for 2004-2010. The flooded area in Upper Lubigi is located mostly within a wetland (or former wetland) area, itself unsuitable for urban development. This being so, flood depth represents no additional constraint to urban development and the simulation of urban growth can be then conceived as an exogenous input to flooding. Alternatively, the absence of flood among the predictors of development could indicate a high level of inertia within poverty bound populations. They stand to loose too much (low cost access to jobs, cheap housing, livelihoods, familiar surroundings and social networks) by leaving a hazardous area, making them uncertain about the benefits of moving to a safer place.

In any case, the integrated model can also be used to explore the feedback between flooding and urban growth which may exist in other sub-catchments. Computationally, the adjustment to algorithm 1 is straightforward: instead of running a dynamic urban growth model (for periods t_0 to t_n) and using its outputs to run the flood model (only for period t_n), both models (UGM and flood) should be run sequentially for each period t (and for a single period, the outputs of the UGM would be the inputs of the flood model). The outputs of this period, t, would then be used as inputs for the subsequent period, $t + 1$.

More interesting are the *conceptual* questions which arise. For example, during one year (the temporal resolution of the urban growth model), up to 365 flood events may occur (the storm represents a daily maximum). Which one should be simulated for each year? The strict answer would be, the event which causes urban agents to change their behavior (which convinces them to not build in flooded areas). The best simplification to operationalize such statement is likely to select the largest rainfall event that occurred during this year. Yet it is also perhaps questionable whether urban growth will be significantly affected by a single flood event of the kind often experienced in Kampala or if a longer term cumulative effect reveals itself after a particularly severe rainy season. Precisely because so large uncertainties remain, applying the modeling approach to past yearly data, and in particular to the feedback between flood and urban growth, is a potentially profitable area of future research.

4.4.5 Urban resilience in the Lubigi catchment

The analysis results reveal collective actions to mitigate flooding, namely past investments in the main drainage channels and the potential of preserving the wetlands, are not sufficient to reduce flood risk in Upper Lubigi. Yet, simultaneously, city life in this part of Kampala continues unimpeded, even in the face of recurring flooding. Could this be evidence of some otherwise unsuspected source of urban resilience?

While governance and communal structures surely contribute to urban resilience, Campanella (2006) has argued that the ultimate source of resilience in a city is its people. However, there is no evidence of a particularly engaged citizenry in Upper Lubigi, on the subject of adaptation to flood risk. Chereni (2016) examined the Bwaise 3 informal settlement, located near the outlet of the Upper Lubigi sub-catchment. He found little evidence of robust social institutions promoting resilience: social networks somewhat influenced the adoption of mitigation measures (58% of surveyed households adopted such measures in response to social influence) but income level, occupation, perception of flood risk or exposure, and even experience of past floods were all uncorrelated to the adoption of household-level mitigation actions.

The mismatch between suffering recurrent floods and failing to proactively address such risk should perhaps be best interpreted as the combination of an acceptance of flood inconvenience and a relatively small damage suffered by each household, since widespread poverty among those most affected means fewer assets exist to be damaged (such characterization is consistent with Kampala as exemplifying a fatalistic culture, in the sense of grid-group culture theory, see Mamadouh, 1999). This subject requires a deeper exploration, which is beyond the scope of this chapter. What is clear from the modeling results is that even efficient land use planning and investments in large infrastructure systems are not enough to reduce flood risk in Upper Lubigi.

4.5 Conclusions

This chapter has summarized the implementation of an integrated land cover change and flood model. The modeling ensemble proved to be operational by successfully simulating prospective land policies and assessing their flood impacts. While not incorporating a feedback between flooding and land cover change – because it was not required by the case study at this point–, the formulation can be easily extended, should the analysis of a specific case require it.

The scenario assessment successfully evaluated a realistic land policy, a defensive landscape strategy such as the stringent protection of the wetlands. This policy was predicted to be ineffective, when compared to trend conditions. Therefore, and in a context of a study area (Upper Lubigi) which already has flooding problems, much more aggressive policies will be required to mitigate future urban growth impacts.

The developed urban growth model has proved capable of replicating existing land trends using a simple, comprehensible approach. The model divides the processes of allocation (where urban development occurs) and the simulation of growth (how much development occurs). Allocation is the result of a weighed summation, which results in transparency on the choice of behavioral assumptions and their formalization into the model. The simulation of growth, in turn, provides flexibility to tackle diverse urban growth conditions. For the current application, $SimDem_i$ is based on randomness, to replicate the dynamics of Upper Lubigi. It may also be applied with development supply based on, for example, land use regulations – such an application was successfully tested for densification scenarios of the city of Kigali, Rwanda (Pérez-Molina et al., 2016). In general, the use of a separate and spatially explicit model of supply may prove useful to a class of cellular automata models (such as Yeh and Li, 2001 or van Vliet et al., 2012) that use a continuous value as the response variable.

The modeling ensemble was deployed in a data scarce environment. The use of remote sensing to develop land cover data models and field measurements of soil characteristics, complemented with digital elevation models, permitted the bridging of this data gap. However, it is likely that considerable uncertainty was introduced into the modeling results. Indeed, further research into error propagation remains a key research need in spatial modeling. Validation of remote sensing-based data models of land cover is an important step to control for the uncertainty introduced into the urban growth and flood modeling results.

Joint modeling of land systems and flood can contribute to improve land use planning in hazardous contexts. The modeling strategy could also be profitably applied to other spatially differentiated hazards, impacted by land cover changes; since it is based on PCRaster – a very flexible environmental modeling software –, the development of other hazard models is feasible. In synthesis, land models can be successfully developed and coupled with natural hazard models, and used to formulate land policies to mitigate their impacts.

5

A Markov Chain Monte Carlo calibration approach for dynamic urban growth models[1]

Abstract

The calibration and validation processes for urban growth models of Kampala, Uganda (2001-2016) and Kigali, Rwanda (2000-2015) are described; calibration is based on the Metropolis-Hastings algorithm to derive transition rules. These transition rules were used to test the hypothesized model structure, a set of spatial factors that are potential determinants of land cover change in Sub-Saharan Africa. The calibrated models were used to simulate urban growth for 2001-2016 in Kampala and 2000-2015 in Kigali. The resulting land cover predictions were characterized through measures of global landscape agreement. Calibration and validation were achieved by comparing the model predictions for an intermediate year (2009/2010) with independent land cover data for that period, as well as the overall trend of landscape evolution. The validation analysis characterizes the results of model development in terms of parameter and data uncertainty. The model distinctly improves simulated results, relative to random parameters and data; further, uncertainty introduced by parameters causes less problems, relative to the prediction, than uncertainty in the spatial determinants of the model.

Keywords: Cellular automata, Calibration, Validation, Land cover change, Urban growth model, Markov Chain Monte Carlo, Metropolis-Hastings, Kampala (Uganda), Kigali (Rwanda)

[1] This chapter is based on: Pérez-Molina et al. (2019b).

5.1 Introduction

Cellular automata models have a long history of successful use in the field of land systems modeling, particularly in urban contexts (Santé et al., 2010). These models conceptualize space as an array of cells. They simulate spatial dynamics by means of transition rules that control how the state of each one of these cells continually changes in response to the state of surrounding cells (Rasmussen and Hamilton, 2012).

Cellular automata models, and land systems models in general, have been primarily used to project future land cover (and land use) patterns (van Vliet et al., 2016). The quality of these results is crucially dependent on the success with which they can *realistically* replicate the fundamental dynamics of the land change process. Land change processes, in particular land cover change, are generally formalized as a spatially explicit relationship between factors representing the human locational preferences of certain activities and the landscape resulting from the transformation these activities cause on the environment.

The process of model development for a specific case is, therefore, centered on model calibration and validation. The problem of calibration is, given a set of input data, to choose the model parameters such that the model outputs reproduce measurements of the modeled phenomenon associated to these inputs (Rykiel, 1996; Refsgaard and Henriksen, 2004); it is a stage of learning about the system from the data. The problem of validation is one of assessment: given a set of model parameters and data inputs, to determine how well do model outputs agree with data about the modeled phenomenon (Rykiel, 1996; Refsgaard and Henriksen, 2004). Both modeling operations share the need for measurements of the coincidence between the model results and independent data measurements of the system's response (Pontius Jr, 2002; Hagen-Zanker and Martens, 2008).

This chapter is organized around the question, how can an urban growth model be calibrated to, simultaneously, characterize the land process dynamics and create realistic projections of trend conditions? To tackle this issue, Bayesian econometrics is used to calibrate an upgraded version of the cellular automata model of urban growth described in chapter 4; specifically, models of urban growth are calibrated for the Kampala metropolitan area for 2001-2016 and for the Kigali metropolitan area for 2000-2015, using a Markov Chain Monte Carlo (MCMC) method – the Metropolis-Hastings[2] algorithm (see a description in Gilks, 1996 as well as subsection 5.3.1), originally proposed in 1953 and generalized in 1970 (Hitchcock, 2003) – to characterize the statistical distributions of model parameters. From these distributions, one may determine the relative importance of proximity, accessibility, and physical factors that determine urban growth in Kampala. Urban growth is then simulated

[2]What is referred to in this chapter as the Metropolis-Hastings algorithm is the random walk approach originally proposed by Metropolis and colleagues; several expositions consider alternative algorithms, such as the Gibbs sampler, as special cases of the broader Metropolis-Hastings framework (Chib and Greenberg, 1995).

using Monte Carlo techniques for the entire period under study, and these simulations are validated by comparing them to independently derived land cover maps of Kampala for 2010 and of Kigali for 2009.

5.2 Literature review

Cellular automata have been widely, and successfully, used to model land use and land cover change processes. General purpose models and software, such as SLEUTH (Chaudhuri and Clarke, 2013) or DINAMICA EGO (Rodrigues and Soares-Filho, 2018) have been adapted and applied to a variety of specific contexts. Indeed, the flexibility permitted by these and by customized modeling approaches that implement cellular automata has resulted in a large variety of approaches (Verburg et al., 2004), although generally cell-based models consider a relatively narrow set of representations of human behaviors (van Schrojenstein Lantman et al., 2011; Geographical Sciences Committee and others, 2014) such as accessibility to urban centralities, agglomeration effects, and physical suitability.

5.2.1 Cellular automata models of land systems calibration and the map comparison problem

A recent review of calibration and validation practices (van Vliet et al., 2016) revealed that statistical analysis and computational methods have become the most common approaches to calibrate land change models. The majority of these statistical analysis cases, as reported in the supplementary data, correspond to regression analysis, which is usually static. Validation practices are less advanced: most applications base their validation on mere locational analysis (as opposed to the global and regional scales of the landscape) and nearly one third do not report any validation (see the supplementary data of van Vliet et al., 2016 for a breakdown of the data).

Why does land use modeling practice lack the calibration and, especially, validation emphasis common to other quantitative modeling domains? A possible answer to this question relates to the complex systems characteristics exhibited by land patterns (Batty, 2009): these emerge from a myriad of interactions between multiple urban agents and amount to regularities in space which are the consequence of these interactions. The complex character of land patterns results in three important particularities for land systems modeling: (1) urban patterns present scaling relations in the response variable and in its determinants (Batty, 2009), because of which parsimonious models are capable of replicating the evolution of urban patterns; (2) extending parsimonious models, e.g. to evaluate the impact a given determinant, poses risks of tautology (of using the scaling relations of the determinants to predict the scaling relations in the response variable), overfitting (an interesting discussion on this subject is developed by Brown et al., 2005, in the context of agent-based models of land), and equifinality (van Vliet et al., 2016); (3) given its character as emergent phenomenon, there is a methodological question of how to describe the land pattern (a central issue for validation and calibration of land models, since they rely on comparison between a predicted pattern and an independently derived map of the pattern; Hagen-Zanker & Mertens, 2008).

The problem of map comparison, and especially extending it beyond location-to-location agreement, is not new and the complications have been acknowledged for many years. Pontius Jr (2002) proposed an assessment to understand differences between two maps as those between location and between quantity (of a land cover category), pointing out as well the need to consider multiple scales. van Vliet et al. (2013) extended the traditional measure of map assessment, the Kappa statistic, by allowing for uncertainty of location in the immediate vicinity of a cell, thus providing a more accurate assessment of the pattern agreement – since cell-to-cell agreement may underestimate the agreement of the overall pattern; they also introduce uncertainty by assessing expected (simulated) rather than observed agreement. An alternative to fuzzy measures is the use of landscape metrics as global measures of landscape structure; these have been implemented, in the context of Bayesian calibration of land system models, by Verstegen et al. (2014).

5.2.2 Bayesian statistics applied to calibration of cellular automata models of land cover

A review was conducted of previous applications of Bayesian methods to the calibration of dynamic spatially explicit land system models. The search was conducted in ISI Web of Knowledge for the period 2010-2018, using the search criteria: (cellular+automata, refined by: TOPIC: ((land* OR urban)) AND TOPIC: (Bayes*); an additional check was performed in Google Scholar, searching for the terms: "cellular automata" AND "metropolis hastings" AND "land". Two calibration techniques were found to have been implemented in dynamic models, namely Sequential Monte Carlo and a Markov Chain Monte Carlo (MCMC) method, the Metropolis-Hastings algorithm.

The search resulted in four cases that applied Bayesian techniques to calibrate cellular automata models. Verstegen et al. (2014, 2016) calibrated an extended cellular automata model of Brazil to explain sugar cane land cover. Their objective was to, firstly, establish the model structure and parameterization (Verstegen et al., 2014) and, secondly, to test the stability in time of these parameters (Verstegen et al., 2016). Rasmussen and Hamilton (2012) created a theoretical model of range expansion, applicable to biological invasion; population changes follow from two parameters which represent short and long distance immigration. They used synthetic data to test a calibration method and recover the parameters. Somodi et al. (2011) modeled the spread of vegetation and of seeds for nine vegetation types, treating transition matrices for missing vegetation maps as parameters and using Bayesian techniques to sample these maps. Mustafa et al. (2017) also calibrated an extended cellular automata model of land cover change for southern Belgium, allocating urban growth based on a weighed summation of spatial factors. The first set of factors, the interaction effects between land uses at distance ranging from one to five cells, were determined by use of Bayesian methods whereas the weights of other static factors was set through a logistic regression model.

Verstegen et al. (2014, 2016) and Rasmussen and Hamilton (2012) applied Sequential Monte Carlo methods – Verstegen et al. (2014, 2016) the Particle Filter and Rasmussen and Hamilton (2012) what they term the Population Monte Carlo Approximate Bayesian Computation. Sequential Monte Carlo relies on propagating forward in time the model states of sampled sets of parameters; for periods when observations of the system are available, the sampling process is updated so the information from the observations is incorporated into subsequent sampling (van Leeuwen, 2009). Mustafa et al. (2017) and Somodi et al. (2011) applied the Metropolis-Hastings algorithm, a common type of MCMC analysis. MCMC methods differ from Sequential Monte Carlo in the sampling scheme: for sequential methods, each sampled instance is independent from all other; in MCMC, these samples are correlated: the sampling scheme creates a Markov chain in which a sampled instance depends on the preceding – concretely, a sampled set of parameters is added to the posterior distribution if it generates a better prediction than a random level, and is rejected otherwise.

A key issue in the application of Bayesian methods to land use/land cover modeling is the definition of how good a prediction is, relative to the observed data. Verstegen et al. (2014) made use of three measures of land patterns to judge this agreement, two global (the number of patches and an edge index, measuring the overall complexity of the patches) and a zonal measure, the focal average within a grid of 150 km squares. Mustafa et al. (2017), on the other hand, used only cell-to-cell location agreement (a local measure). The results of these two cases demonstrate the sphere over which the calibration yields satisfactory results.

Mustafa et al. (2017) report a 32.75% cell-to-cell location agreement between their prediction and the land use target map over a ten year period for cells undergoing change, in line with similar cellular automata models of urban growth. Most case studies of cellular models reporting accuracy percentages cited by van Vliet et al. (2016) report overall percentages of over 80%; only Hansen (2014), while presenting an overall cell-to-cell agreement of 95%, also reports a 23% agreement when considering only the change cells. While this figure is hardly impressive, cell-to-cell location of change cells is probably too fine a measure to properly judge the results of a land use model – indeed, it has been argued (Hagen-Zanker and Martens, 2008; van Vliet et al., 2013) that calibration and validation of urban growth models should aim to reproduce the pattern, overall and regionally rather than merely at the cell level – , and it is remarkable that such a measure was able to successfully guide the Metropolis-Hastings algorithm.

Verstegen et al. (2014, 2016) provide greater methodological detail on the model development process. From a set of potential spatial factors, chosen by expert knowledge, their method is capable of discarding those that are irrelevant as well as determining the weight of spatial factors that are relevant to the land cover change process. Relative to a reference model that does not apply the Particle Filter, i.e. only the Monte Carlo simulations, the Particle Filter reduces substantially the RMSE of the three measures of land pattern; further, the median value of each of these measures is better approximated by the application of the Particle Filter. Verstegen et al. (2014) also note that, as the simulation progresses from the time period in which the Particle Filter is applied, the model performance decreases; in other words, they can generate reliable information about the land use system only in the short-run; indeed, Verstegen et al. (2016) document how the model parameters change yearly in response to new land information.

5.3 Methods: Markov Chain Monte Carlo for calibrating urban growth models

5.3.1 Application of Bayesian statistics with MCMC

The process of Bayesian inference begins with an initial probability statement about the parameters *before* observing the data, the so-called prior distribution, which is modified based on a combination of this prior distribution and the data to produce a posterior distribution (Congdon, 2006). The relationship between these three elements can be expressed by the Bayes theorem:

$$p\left(\theta|y\right) \propto p\left(y|\theta\right) \cdot p\left(\theta\right) \tag{5.1}$$

with θ the set of parameters to be estimated conditional on data y (the state of the system and its determinants), $p\left(.|.\right)$ a conditional probability distribution, and $p(.)$ a marginal probability distribution. The prior knowledge about the parameters is represented by $p\left(\theta\right)$, the likelihood by $p\left(y|\theta\right)$, and the posterior distribution is $p\left(\theta|y\right)$ (Congdon, 2006).

As the exact estimation of the likelihoods may involve challenging integration of expressions, it is common practice to use Monte Carlo simulations. A number of N independent values of θ are drawn from the parameter distribution, $p\left(\theta\right)$. The empirical distribution of the sample $\{\theta^1, ..., \theta^N\}$ constitutes and approximation to the posterior distribution; this approximation improves as N becomes larger (Hoff, 2009).

Different methods can be used to choose the θ_i values. The implemented MCMC-based algorithm, the Metropolis-Hastings, operates as follows (Gilks, 1996, p. 84):

- Initialize $t = 1$, draw a parameter value ϑ_1, calculate $p\left(\vartheta_1\right) \cdot p\left(y|\vartheta_1\right)$
- For iteration t:
  - Increment t

– Sample a point ϑ_t from $h(.)$
– Sample a $\mathcal{U}(0,1)$ random variable U
– If $U \leq min[1, [p(\vartheta_t) \cdot p(y|\vartheta_t)] / [p(\vartheta_{t-1}) \cdot p(y|\vartheta_{t-1})]]$ accept ϑ_t
Else, choose $\vartheta_t = \vartheta_{t-1}$

where $h(.)$ is called the proposal function. The following form for proposal functions was used: $\vartheta_t = \vartheta_{t-1} + \mathcal{N}(0, s)$, and s was chosen so that the acceptance rate (the total number of iterations into the amount of iterations for which $U \leq min[1, [p(\vartheta_t) \cdot p(y|\vartheta_t)] / [p(\vartheta_{t-1}) \cdot p(y|\vartheta_{t-1})]]$ is true) is approximately 0.20.

The Metropolis-Hastings is a less popular alternative, among MCMC algorithms, to the Gibbs sampler: this is because the Gibbs sampler does not require a proposal function; however, it does imply the need to analytically derive the posterior distributions (see discussions on their implementation in Chib & Greenberg, 1995 for the Metropolis-Hastings algorithm, specifically the random walk, and Gilks et al., 1994 on the Gibbs sampler). Given an implementation with spatial data, it is very difficult to construct such posterior distributions for the parameters from a theoretical standpoint alone, a problem compounded in the specific case of land cover data by the relatively large amount of noise in the data (in turn an unavoidable consequence of the complexity of the process leading to the human land pattern). Because of this, the Metropolis-Hastings algorithm was implemented.

5.3.2 Urban growth model calibration and validation

Urban growth model specification

The calibrated urban growth model is an extension of the model for the Upper Lubigi sub-catchment (Pérez-Molina et al., 2017). Figure 5.1 schematizes the elements of the model and their relationships. Key features of this model are: the land cover and the amount of urban growth are continuous values (fraction of built-up land cover in a cell); in the case of urban growth, this means the existence of a limit to the resulting (after growth has occurred) built-up fraction, in turn limiting the maximum possible growth (increase of built-up fraction) in a different way for each cell. Relative to the model of Upper Lubigi, two extensions have been made: (1) the model was spatially extended, from the $27.8 km^2$ of Upper Lubigi, to include the entire Kampala metropolitan area, $2650 km^2$, and it also adopts a metropolitan area extent for Kigali (an extent of $547.7 km^2$) and (2) the use of principal components was introduced to deal with correlation problems between spatial factors, specifically three factors related to accessibility (travel time to CBD and distance to main roads) and agglomeration (built-up neighborhood factor).

The model estimates a suitability map that synthesized six spatial factors (maps), one of them dynamic. Each factor was normalized by dividing the range of values into either (a) each cell value minus the minimum of the map, if the relation between the spatial factor and urban

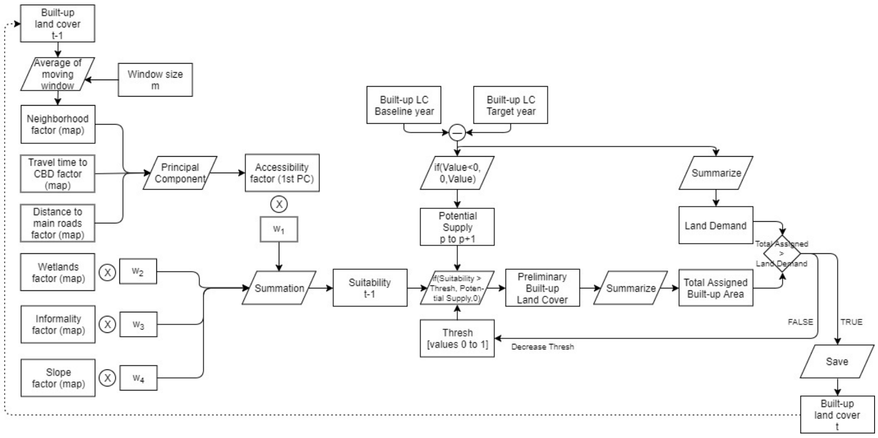


Figure 5.1 Cellular automata urban growth model

growth is direct, or (b) the maximum of the map minus each cell value, if the relation to urban growth is inverse.

Three factors are highly correlated to each other: (1) a neighborhood factor, the average percentage of built-up from the preceding period surrounding each cell (within a moving window; the size of this window, m, is a model parameter), (2) an estimate of travel time to the CBD which incorporates the effect of main roads, local roads, and walking, and (3) the Euclidean distance to main roads. Each of these factors was normalized; the neighborhood effect is thought to be directly related to urban growth and both travel time to CBD and distance to main roads, to be inversely related. To improve the efficiency of the calibration algorithms, the first principal component of these three normalized factors – denoted as the accessibility factor – was included as a determinant of urban growth directly related to urban growth.

Three additional factors, slope percentage, informal settlement locations, and wetland locations, were also incorporated into the model as determinants. Location maps were coded with values of 1.00 to indicate the presence of an element and of 0.00 to indicate absence; they are, therefore, inversely related to urban growth. Slope is also inversely related to urban growth.

The cellular automata model follows the general form synthesized by Batty (2009) and summarized in equations 5.2 and 5.3. Equation 5.2 is called the diffusion equation by Batty (2009). Equation 5.3 corresponds to the transition rule that controls the cellular automata's evolution.

$$BUfr_{i,t} = BUfr_{i,t-1} + \Delta BUfr_{i,t} \tag{5.2}$$

$$\Delta BUfr_{i,t} = \begin{cases} PotSupply_{i,t}, & \text{if } suitability_{i,t} > threshold_t. \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

The suitability map is a weighed summation of the four normalized factors we have described and, by construction, it is directly related to urban growth. The suitability value of each cell is indirectly determined by the built-up fraction (in the preceding time step) of the surrounding cells, where surrounding is defined by the moving window of size m. Hence the argument here adopted that this model is an extended cellular automata model (White, 1998), for which the state of neighboring locations are combined with other spatial factors to determine the state of each cell.

Equations 5.2 and 5.3 are used to implement a constrained cellular automata. For any given period, the $PotSupply_{i,t}$ map is reclassified based on an arbitrary $threshold_t$ of suitability: the cells with high suitability retain the value whereas the cells with low suitability are set to 0.00. The algorithm starts with very large $threshold_t$ values (near 1.00) and works its way towards smaller ones by small increments, stopping when $\sum \Delta BUfr_{it} > Land\ Demand_t$, which are approximately equal if the difference between iterated threshold values is very small.

The potential supply of new urban development, $PotSupply_{i,t}$ is taken as fixed, as described in subsection *Calibration and validation*, since this chapter reports on a calibration exercise. It is important to note that this element poses the general problem of a spatially explicit representation of potential built-up increase, a problem that may have different answers (for example, see on Upper Lubigi, Pérez-Molina et al., 2017 or on the Rwampara wetland of Kigali, Pérez-Molina et al., 2016). This flexibility is very important, especially when simulating forward in time from the present into the future, for which the spatial distribution of the land cover at the target year is unknown.

Landscape indices to evaluate cellular automata model predictions

To estimate the likelihood of a state (built-up land cover), the approach proposed by Verstegen and colleagues was followed: three spatial metrics, selected "based on their complementarity (global vs. regional, configuration vs. composition[)]" (Verstegen et al., 2014, 126). These are: one focal measure, fraction of the land cover in a larger zone (estimated as follows: the study area was divided into squares of 1.00 km to a side; the average built-up fraction per square was estimated – input land cover maps having a spatial resolution of 90 m), and two global measures – total number of patches, a measure of landscape complexity, and a landscape index defined as:

$$q_t = e_t / \left(4 \cdot \sqrt{area_t}\right) \tag{5.4}$$

with e_t the total length of the edge of all urban patches in t, $area_t$ the sum of area for all urban patches in t. q_t compares the perimeter of actual patches with a "minimum" sized patch, a maximally aggregated patch attained if all urban patches were to be grouped into a single square patch.

Since the global measures require binary data inputs, the results were re-classified: a cell was considered urban if its built-up fraction exceeded 0.35. The focal measure, however, was estimated using directly the built-up fraction estimates. For the calculation of probability, a single summary measure of the focal statistics is necessary; the average across the zones was chosen and this average was used to estimate the probability (thus the focal measure becomes global by methodological necessity).

The probability of a given state (end-result simulation for the target year) is given by (van Leeuwen, 2009):

$$p\left(\mathbf{o_t}|\mathbf{s_t}\right) = A \cdot e^{-1/2 \cdot \left(\mathbf{o_t}-\mathbf{s_t}\right)^T \times \mathbf{R_t}^{-1} \times \left(\mathbf{o_t}-\mathbf{s_t}\right)} \tag{5.5}$$

where $\mathbf{o_t}$ is the observed data, $\mathbf{s_t}$ the modeled system state, $p\left(\mathbf{o_t}|\mathbf{s_t}\right)$ is the probability that the observed values occurred, given the modeled state variables, and $\mathbf{R_t}$ the covariance matrix of the observation error.

The calculation of probability requires an estimate of the covariance matrix of the error term in the observed data (term $\mathbf{R_t}$ in equation 5.5). Since the observed data is a single instance (a map of a given period), the procedure outlined by Verstegen et al. (2014) was followed to simulate this covariance matrix:

- The built-up percentage land cover map was transformed into a vector map of points. A sample of 10% of these points was randomly selected.
- The gstat tool (Pebesma and Wesseling, 1998) was used to create 100 conditional simulations, based on a semi-variogram model of this sample.
- For each of the simulated instances, the three spatial metrics were calculated. The error covariance matrix was derived from this dataset.

Calibration and validation

Input data of the land cover for the urban growth model, for each case study, consists of built-up land cover fractions for three periods: a baseline year (2001 for Kampala, 2000 for Kigali), a target year (2016 for Kampala, 2015 for Kigali), and an intermediate year (2010 for Kampala, 2009 for Kigali). To calibrate the urban growth model, the baseline year was projected into the target year (i.e. the 2001-2016 period as simulated for Kampala, and 2000-2015 for Kigali), using urban growth between circa 2000 and circa 2015 as the potential supply. This map was generated through map algebra, by subtracting the baseline year land cover map from the target year map.

The objective of the model's calibration process is to derive distributions of the parameters of the model, which are: the weights of the different spatial factors (four weights, one each for accessibility, wetlands, informality, and slope) and the neighborhood window size. Since there is no prior knowledge on the weight of each factor (other than the assumption that, because of their normalization, they are all directly related to urban growth, i.e. they are positive), a prior uniform distribution was adopted. As for the neighborhood size, the specification chosen follows Verstegen et al. (2014) in defining a log-normal distribution with a parameter such that the median value is equal to three cells (270 m) and the mean, to five cells (450 m), because the window size is assumed to be relatively small; however, because the model becomes undefined if $m < 90$ (one cell), any sampled value under 90 m is rejected. Thus, equation 5.6 summarizes the prior parameter distributions that are inputs to the model calibration algorithm:

$$\begin{aligned} w_{Accessibility} &\sim \mathcal{U}\,(0.00, 2.00) \\ w_{Wetlands} &\sim \mathcal{U}\,(0.00, 2.00) \\ w_{Informality} &\sim \mathcal{U}\,(0.00, 2.00) \\ w_{Slope} &\sim \mathcal{U}\,(0.00, 2.00) \\ m &\sim Lognormal\,(5.6, 1.0) \end{aligned} \tag{5.6}$$

The calibration itself was achieved by comparing the prediction for the intermediate year (2009 or 2010) with the independently derived land cover map corresponding to that year. Two chains were sampled for each case study. Each chain for Kampala was generated with 16500 instances, of which the first 1500 were dropped as burn-in. The full sample included, therefore, 30000 iterations. The full sample was thinned, taking one out of each ten iterations to reduce autocorrelation problems. The final sample consisted of 3000 instances. For Kigali, an equivalent process was followed: two chains of 8000 instances were sampled, with the first 500 iterations dropped as burn-in; the full sample, of 15000 instances, was also thinned by taking one out of ten iterations, resulting in a final sample of 1500 instances.

Sample size was defined such that the Gelman-Rubin statistic indicated convergence. The statistic is defined by comparing the within (a given chain) variance of a parameter with the between (different chains, in this case 2) variance of said parameter. Specifically, if for a parameter the ratio of (a) pooled variance (a weighed summation of within and between variance) and (b) the within variance, is close to 1.00, convergence has been achieved. When all parameters present such ratios, the model has achieved convergence. See Gelman and Rubin (1992) for a full description of the test, as well as Hartig et al. (2018) for documentation on its implementation.

The resulting chains (samples) were analyzed using the packages *BayesianTools* (Hartig et al., 2018) and *coda* (Plummer et al., 2006) from *R* (R Core Team, 2017), as well as general statistical tools of *R* itself.

However, this approach leaves no independent data for validation (i.e. data that has not been used for the model's calibration). In consequence, while land cover was predicted between 2000/2001 and 2015/2016 and in the entire extent, for calibration purposes the landscape indices discussed in subsection *Landscape indices to evaluate cellular automata*

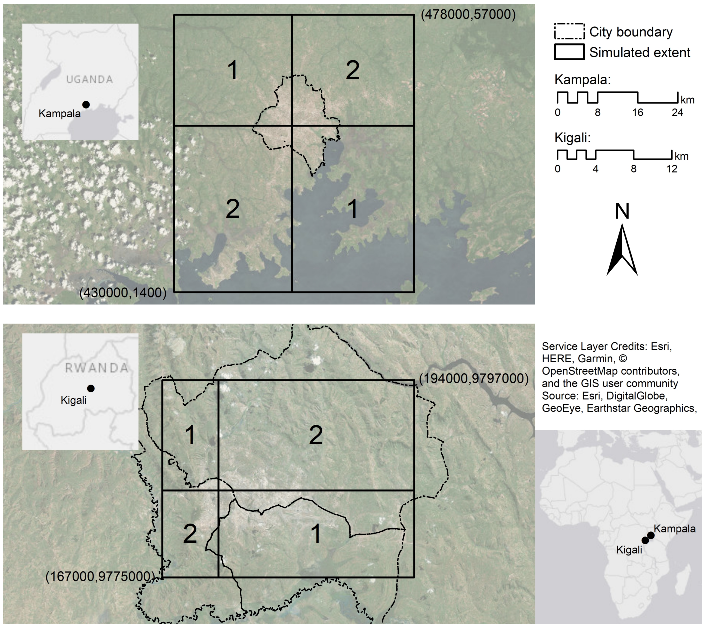


Figure 5.2 Study area: simulated extents of Kampala, Uganda and Kigali, Rwanda

model predictions were only calculated for the quadrants marked with 1 in figure 5.2 (the area within quadrants 2 was set to null data); conversely, for validation, the predictions of the area of quadrants 2 was used, and the area of quadrants 1 were set to null data. The quadrants were delimited by north-south and east-west axis passing through the CBD.

Validation was achieved through Monte Carlo simulations. For each city, 5000 instances were simulated starting with the baseline (2000/2001) up to the target year (2015/2016), and using the same potential supply map as the calibration process (target map minus baseline map). Similarly to calibration, the success of the model was judged based on a comparison of the landscape metrics of the predicted intermediate year (2009/2010) with those of the independently derived land cover for that year (and limited to the areas of quadrant 2).

Three alternative versions of the model were designed to compare the uncertainty introduced into the model by spatial data vs. model parameters; further, these alternative versions permit the exploration of possible equifinality in the data:

1. The first validation model is termed *Random parameters*; it simulated the evolution in the same way as the *Parameterized model*, except the parameters were drawn from the prior distributions, equation 5.6, not from the resulting posteriors (i.e. the calibrated results).

2. The second validation model is a constrained cellular automata

model without any ancillary information; the neighborhood factor is the sole driver of this model and its window size parameter is adopted from the calibration results.

3. The third to fifth validation models rearrange, in a spatially random pattern, the normalized spatial factors: the *Full random* model randomizes all four spatial factors, the *Accessibility random* model only the travel time to CBD and the Euclidean distance to nearest major road, and the *Physical random*, only the slope, informality, and wetlands factors.

5.3.3 Study area

Kampala, Uganda

The extent of area simulated in Kampala, shaped like a rectangle, goes beyond the limits of the Kampala Capital City Authority, i.e. the city proper, to include the urban fringe where much of the recent expansion has taken place (see figure 5.2). To the south, it is bounded by Lake Victoria. The rectangle's limits are, to the west, UTM 36N coordinate 430000 m, to the east, coordinate 478000 m, and to the north, coordinate 57000 m. These limits encompass all urban area existing in 2016 plus the peri-urban interface and the surrounding rural area, in which future urban growth will mostly occur.

Kigali, Rwanda

The administrative boundaries of Kigali correspond to a province of Rwanda and, because of this, exceed the limits of the urban area proper. Furthermore, Mount Kigali poses a physical barrier which blocks the city's westward expansion. The extent simulated was selected to encompass the urban footprint of the city (in 2015) and a large enough area to accommodate Kigali's expansion (see figure 5.2). A rectangle bounded by the (UTM 36S) coordinates 9797000 m north, 9775000 m south, 194000 east, and 167000 m west, defines the extent; the CBD is located towards the west of this area, since most of the expected expansion of Kigali will most likely occur towards the east due to Mount Kigali. The land cover models of Kigali were described in chapter 2.

5.4 Results and discussion

5.4.1 Results from calibration of urban growth model

The results of parameter posterior distributions for Kampala are shown in figure 5.3. Evidence of convergence was found in the decreasing autocorrelation (after thinning) and, generally, trace plots ranging over the entire parameter space. The Gelman-Rubin diagnostics for each factor, reported in table 5.1, are all below the 1.1 threshold, indicating convergence was achieved. Some problems were detected, such as the

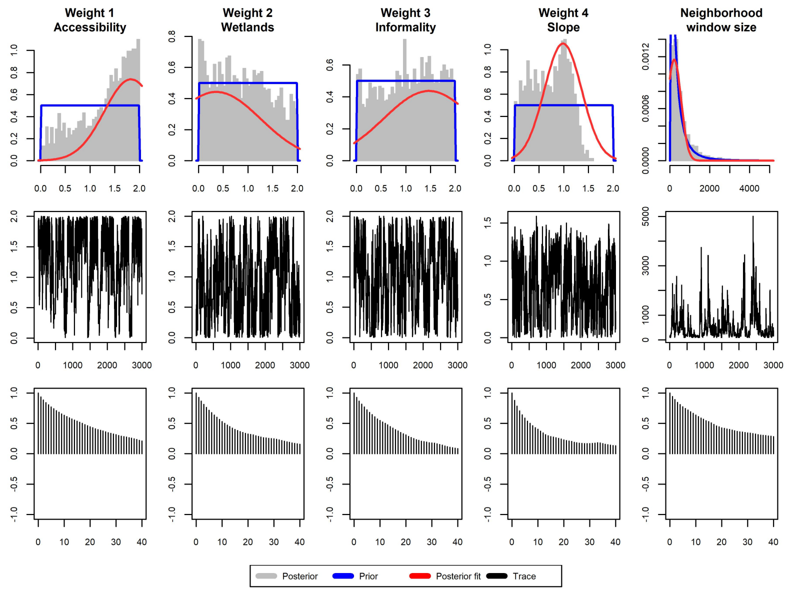


Figure 5.3 Parameter prior and posterior distributions for Kampala. Calibration results. Top: posterior distribution histograms. Middle: trace plots. Bottom: autocorrelation analysis

weight factor for slope not having been sampled for the larger possible weights.

Table 5.1 Gelman-Rubin diagnostics for convergence

Parameter	Kampala model	Kigali model
$w_{Accessibility}$	1.010	1.011
$w_{Wetlands}$	1.000	1.078
$w_{Informality}$	1.030	1.010
w_{Slope}	1.050	1.007
m	1.100	0.999

Two histograms produce reasonably clear results: those of the accessibility and slope parameters. The wetlands and informality parameters' histograms, on the other hand, show a less defined peak, although they also appear skewed: the wetlands histogram towards the lower values, the informality histogram towards the larger values. The neighborhood window size posterior distribution resembles the prior, suggesting little signal has been achieved – and also that the window size does not greatly affect the predictions of the model.

Mean values for the posterior distributions were selected by multiplying the mean values of the thinned sample of each weight times the following factors: 1.4 (for accessibility), 0.4 (for wetlands), 1.4 (for informality), 1.4 (for slope), 0.4 (for the neighborhood window size). Standard deviations for the posterior distribution, in turn, were estimated as those of the sample times: 1.0 (for accessibility), 1.6 (for wetlands), 1.6 (for informality), 1.0 (for slope), 0.6 (for the neighborhood window size). These numbers, reported in equation 5.7, were selected such that the posterior fitted curves, which are all normal distributions, correspond to the histogram of the sample by visual assessment.

$$\begin{aligned}
w_{Accessibility} &\sim \mathcal{N}\,(1.83, 0.54)\\
w_{Wetlands} &\sim \mathcal{N}\,(0.36, 0.90)\\
w_{Informality} &\sim \mathcal{N}\,(1.47, 0.91)\\
w_{Slope} &\sim \mathcal{N}\,(0.98, 0.38)\\
m &\sim \mathcal{N}\,(222.9, 341.7)
\end{aligned} \tag{5.7}$$

Two histograms, of the parameters with better defined peaks (accessibility and slope), were set with standard deviations equal to those of the sample. The two less well defined peaks (of wetlands and informality) were fit with standard deviations that are larger than those of the sample, and partly because of this, such fitted distributions explain less of the sample.

The results of parameter posterior distributions for Kigali are shown in figure 5.4. The model results were better, in formal terms, than those of Kampala: autocorrelation decreases more rapidly, trace plots show a better variation over the parameter space, and (with the exception of the wetlands factor) histograms show clear signals. The Gelman-Rubin diagnostics for each factor, reported in table 5.1, are like in Kampala all below the 1.1 threshold, indicating convergence was achieved. Kigali is a smaller city than Kampala and it has also been more structured, both by a more irregular terrain and a better land use regulation system (Goodfellow, 2013a). It is, therefore, easier to model (relative to the "messier" situation of Kampala) since variation should be expected to be more systematic and its scope is smaller.

In the case of the model results for Kigali, four of the five histograms produce clear signals. The only ambiguous parameter, like the most ambiguous of Kampala, is the wetlands factor. If any, this posterior distribution is slightly skewed towards the larger values; in consequence, it was modeled with a normal distribution by increasing the mean of the data 1.2 times and increasing the standard deviation, 1.8 times – the largest increase of all the fitted factors for either model. The mean values of the fitted posterior distributions for the other factors were determined by multiplying the mean values of the sampled data times: 1.2 (for accessibility), 0.4 (for informality), 1.4 (for slope), 1.0 (for neighborhood window size), and the standard deviations, times: 1.2 (for accessibility), 1.0 (for informality), 1.0 (for slope), 0.15 (for neighborhood window size), seeking in all cases, as in the case of Kampala, a curve judged visually to fit the histogram. The resulting parameters are summarized in equation

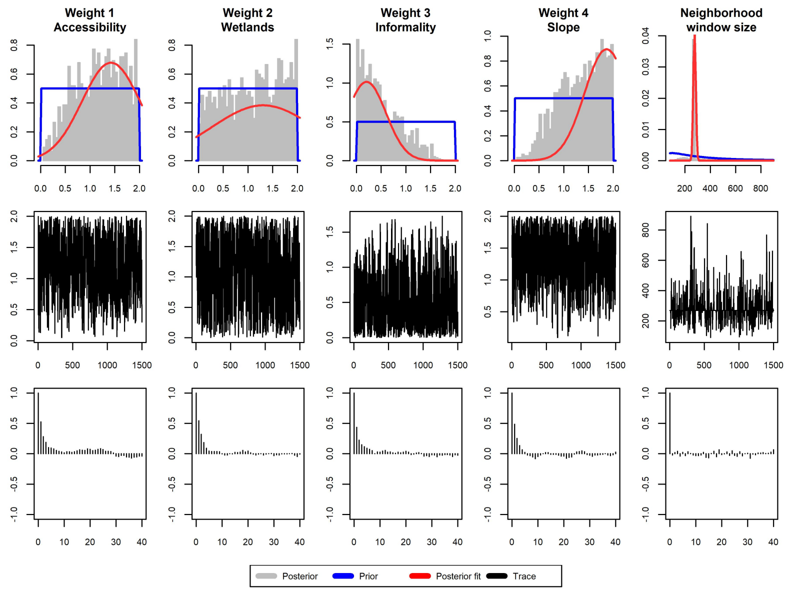


Figure 5.4 Parameter prior and posterior distributions for Kigali. Calibration results. Top: posterior distribution histograms. Middle: trace plots. Bottom: autocorrelation analysis

5.8.

$$
\begin{aligned}
w_{Accessibility} &\sim \mathcal{N}\,(1.42, 0.59)\\
w_{Wetlands} &\sim \mathcal{N}\,(1.31, 1.04)\\
w_{Informality} &\sim \mathcal{N}\,(0.20, 0.39)\\
w_{Slope} &\sim \mathcal{N}\,(1.87, 0.45)\\
m &\sim \mathcal{N}\,(275.6, 9.7)
\end{aligned} \tag{5.8}
$$

When comparing the results of the metropolitan areas, clearly accessibility and slope are the most important factors. This is unsurprising: accessibility is, according to the theory (of urban land rent formation, specifically the Alonso-Mills-Muth model, see Glaeser, 2008 and Brueckner, 1987; the link between land rent and its associated urban location choice by urban agents, on the one hand, and land use/land cover patterns on the other has been discussed in early spatial statistical modeling, e.g. Chomitz and Gray, 1996; on the relation between land use intensity and urban land rent formation, see the model by Brueckner, 1983; typical determinants and processes underlying them for cell-based urban growth models have been discussed by van Schrojenstein Lantman et al, 2011), the central factor in determining urban location. The specific context of the two case studies, on the other hand, already suggested terrain to be a central determinant. Kampala, for example, owes its current location to the XIX century custom of the kings of Buganda of choosing a hilltop for their capital (which they would move every few years within a relatively small area) and their tight control of foreigners within their kingdom, forcing them to live in the capital (Southall and Gutkind, 1957). The terrain in Kigali is, if anything, even more important (as witnessed by the greater mean value of the fitted parameter posterior distribution for Kigali relative to Kampala) because of the very large variation present in Kigali's landscape.

Informality and wetlands are weaker explanations of urban growth patterns, although the signal of the wetlands parameter is likely better for Kampala than for Kigali (i.e, dispersion of the sampled parameter's histogram is smaller) and for informality, the reverse is true.

No definitive explanation was found for the weakness (high level of dispersion of the sampled parameter) of the wetland factor's weight. There is evidence of some wetlands of Kampala having been occupied by urban development, especially in the centrally located wetland of Nakivubo; however, it is also true that the complex land tenure system could act in constraining expansion into some wetland areas (Mabikke, 2016). In the case of Kigali, both the better land use regulation and the still slower growth of the city (leading to less pressure from demand on land for urban development) explain why wetlands are relatively free of built-up land cover. Yet the results also show some wetlands (often centrally located) are attractive for urban growth. Indeed, wetland locations are more important determinants of urban growth in Kigali than in Kampala: for the weight of wetlands location, the mean of the fitted posterior distribution is greater for Kigali than for Kampala (although this posterior distribution of Kigali also has a larger dispersion).

The lesser role of informality could be explained by the better land use regulation system of Kigali (see Goodfellow, 2013, for a comparison of both systems): not only is the dispersion of the sampled parameter's histogram lower than for Kampala, the central tendency measure is also smaller than for Kampala (this factor is less important for Kigali than for Kampala). The city of Kigali is generally considered to be more effective in controlling informal development than the city of Kampala; our results are consistent with this view.

5.4.2 Validation and verification of urban growth model

The landscape measures for the validation simulations, using the calibrated model (*Parameterized model*) and the alternative versions of both cities are reported in figures 5.6 and 5.7. As noted in the methodology, these results correspond only to the area of quadrants 2 (figure 5.2).

While the definition of the landscape metrics was discussed at length, their interpretation becomes at this point crucial to the assessment. The *Zonal average* is the average of built-up fraction across all zones; because it is an average of focal averages, it represents built-up land cover intensity: it is a measure of how much more (or less) intensification of land does the model assume has occurred, relative to the data. Because cellular automata models rely greatly on the neighborhood factor, one should expect the model to predict more intensification than the independent calibration data.

The *Number of patches* is a measure of how fragmented the landscape is; more patches equal a more discontinuous urban fabric. The *Edge Index* measures the complexity of these patches and, in this sense, is also consistent with discontinuous development. However, these two measures correspond to “urban” areas in the restricted sense of places with high degree of built-up land cover (built-up fraction over 0.35). Thus, while both cities (and especially Kampala) have been sprawling, it is also possible these landscape metrics do not fully reflect such process because of its relatively low built-up land cover fraction.

Table 5.2 Validation results: median results from *Parameterized model* simulations and metrics for land cover map corresponding to quadrants 2, year 2009/2010

	Model prediction	Land cover map
Kampala		
Zonal average	0.0723	0.0731
Number of patches	1636	1400
Edge Index	52.5	43.7
Kigali		
Zonal average	0.0596	0.0355
Number of patches	154	118
Edge Index	14.2	13.1

Table 5.2 presents the parameterized model’s predictions for the intermediate year and the metrics derived form the land cover maps. The level of intensification predicted for Kampala (*Zonal average*) is essentially the same as that of the validation data; however, for Kigali, the model assumes substantially more intensification than what was detected in the land cover maps. Relative to *Number of patches* and *Edge Index*, the model produces predictions of a more complex landscape than what was detected from land cover maps, both for Kampala and for Kigali.

The differences in Kigali can be explained by pointing out the difference between the urban growth for 2000-2009 (validation) and for 2000-2015 (calibration), see figure 5.5. When examining the latter, one can describe the process as the original large patches that comprised the urban core of Kigali in 2000 expanding outwards, towards the south and north: thus a relatively compact expansion. However, the urban growth of 2000-2009 saw the emergence of a relatively large and non-contiguous (to the 2000 core) patch of built-up land cover. Events actually unfolded, then, by a first stage of greenfield expansion in 2000-2009 and a later infill development stage in 2009-2015. This is consistent with the policy context, as the land use plans were implemented circa 2009 (Goodfellow, 2013a), which could account (jointly with terrain effects) for the change of trajectory. More importantly, though, since calibration used the 2000-

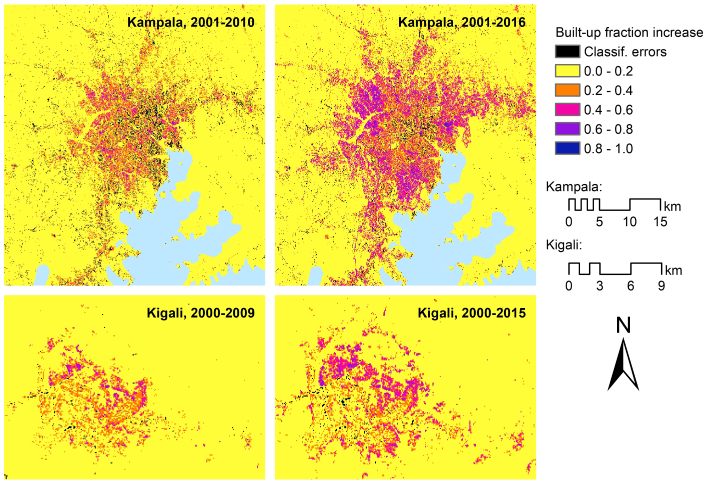


Figure 5.5 Urban growth patterns of Kampala and Kigali

2015 data and validation, the 2000-2009 data, the model has assumed whatever reasons caused the more compact development to have been uniformly influential in time but they had not yet come into play during the validation period.

The case of Kampala (see figure 5.5) has a similar explanation (fundamental changes in the trend of urban development) but with no exogenous cause. Rather, the trend itself evolved spatially: during 2001-2010, urban growth occurred in a typically discontinuous horizontal expansion at low intensity; much area was developed but at low built-up fractions. Therefore, the complexity added to the overall landscape was not detected by the landscape metrics. Contrast this to 2010-2016, which saw simultaneously the intensification of what had been greenfield development in the preceding period and new greenfield development but exhibiting larger built-up fractions. These trends, likely explained by continuing population growth and the absence of effective land use controls, did cause a noticeably more complex simulation of the landscape result. Since the former condition describes the data used for validation and the latter, the calibration, the differences observed in table 5.2 can be thus explained.

In this context, it is useful to bring into the discussion the observation of Batty (2009), when comparing urban growth trends of London and Las Vegas. He observed they were visually very similar, as they both embodied the scaling relations which give predictive strength to cellular automata. However, he noted, the technologies and patterns of human settlement behind these trends were very different and indeed changed substantially in time. Regarding the case studies, it is believed the long run manifestations of the scaling relations are not yet apparent (meaning the trends are still changing), given the limited time extent of our modeling exercise. This affords the opportunity to discuss the trends and their impact on modeling.

This is not to suggest, however, that the limitations of the developed model undermine its usefulness. Indeed, in figures 5.6 (for Kampala) and 5.7 (for Kigali) the 95% credible interval of each landscape metric is compared across different simulations. But for the *Constrained CA* version, the discussion of overall landscape complexity holds: models tend to predict more complex landscapes because of the difference between calibration and validation data periods.

As noted, the *Parameterized model* uses the fully calibrated model and it is, in both cases, the best prediction for 2010. They manifest the smallest bias, regarding the over-complex prediction. The second best prediction corresponds to *Random parameterization*, which draws the parameters from uniform distributions rather than from the calibrated posteriors. The credible intervals of the *Parameterized model* and *Random parameterization* are very similar but the median value is better for the *Parameterized model*.

The impact of spatially randomizing spatial factors (models *Full random*, *Accessibility random*, and *Physical random*) is much larger. The evident conclusion is input data is determinant when generating the prediction of the model and, if improvements are to be prioritized in this type of model, the quality of input data is a more fertile area than parameterization. Furthermore, the randomization of physical factors (informal settlements and wetlands locations, slope) causes more problems than the randomization of accessibility: for *Full random* and *Physical random*, the predictions of both Kampala and Kigali very quickly rise in complexity, reach a maximum and then descend, although not quite reaching the 2015/2016 target year metrics. For *Accessibility random*, which randomizes only Euclidean distance to nearest main road and travel time to CBD, the trend is similar to that of *Parameterized model* and of *Random parameters*.

Two additional comments on the randomization of spatial factors are necessary. In general, they predict much narrower 95% credible intervals, suggesting a strong equifinality in the land cover data that are used in the simulation (although further analysis of this issue could be warranted). The exception is the *Accessibility random* of Kigali, which leads to a very similar result to *Parameterized model* and to *Random parameters*. Secondly, and very clearly, randomizing the accessibility factors has a more limited effect on the predictions: this is because the neighborhood factor is not randomized and it is combined into a single factor with the Euclidean distance to nearest main road and travel time to CBD factors. Apparently, therefore, the neighborhood effect is capable of correcting for very large problems with the input datasets.

Yet a model based purely on the neighborhood factor, such as *Constrained CA*, fares the worst of all. What is shown in figures 5.6 and 5.7 is a model that has not converged. This can be seen in how the credible interval of the *Zonal average* expands towards lower fractions in the later periods of both simulations (of Kampala and of Kigali). The variation of complexity is also erratic: it decreases substantially, since agglomeration

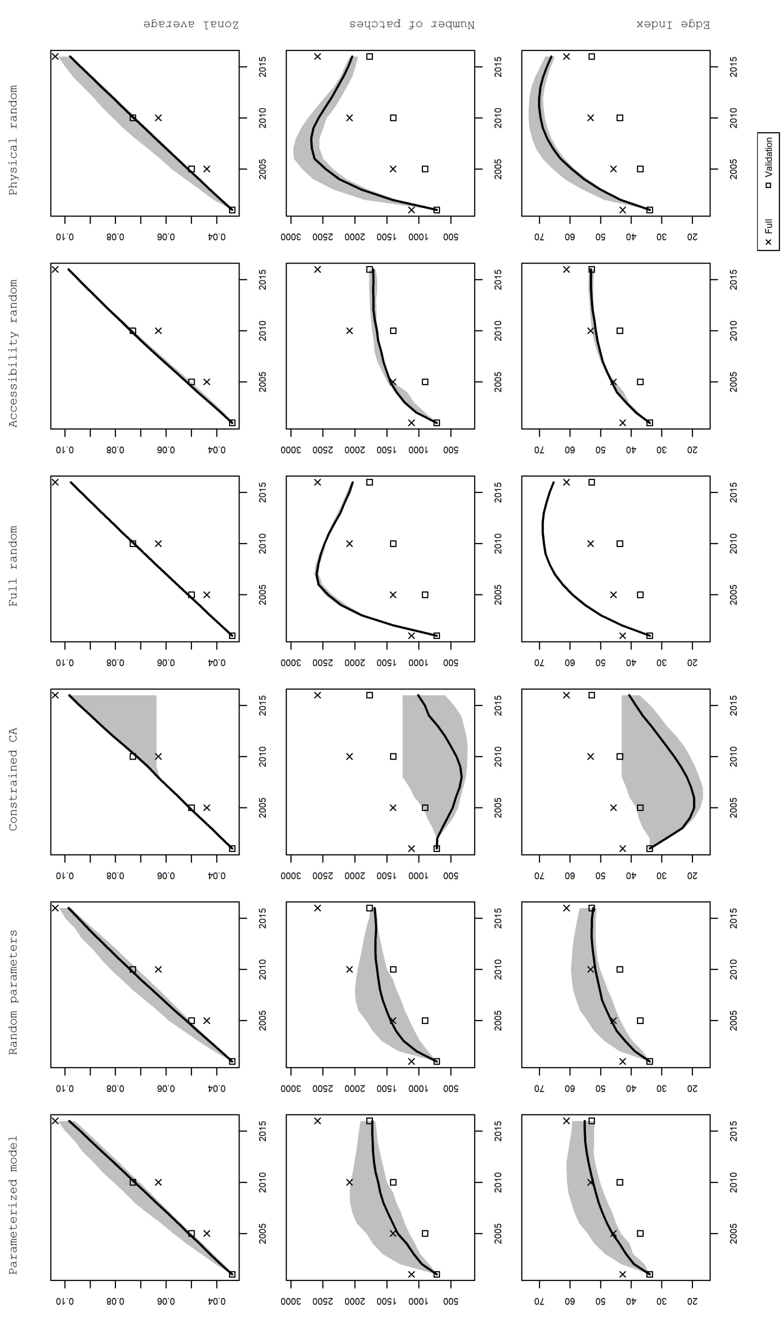


Figure 5.6 Comparison of modelled median (solid line), 95 credible interval (grey area) and observations (points) of the spatial metrics for Kampala

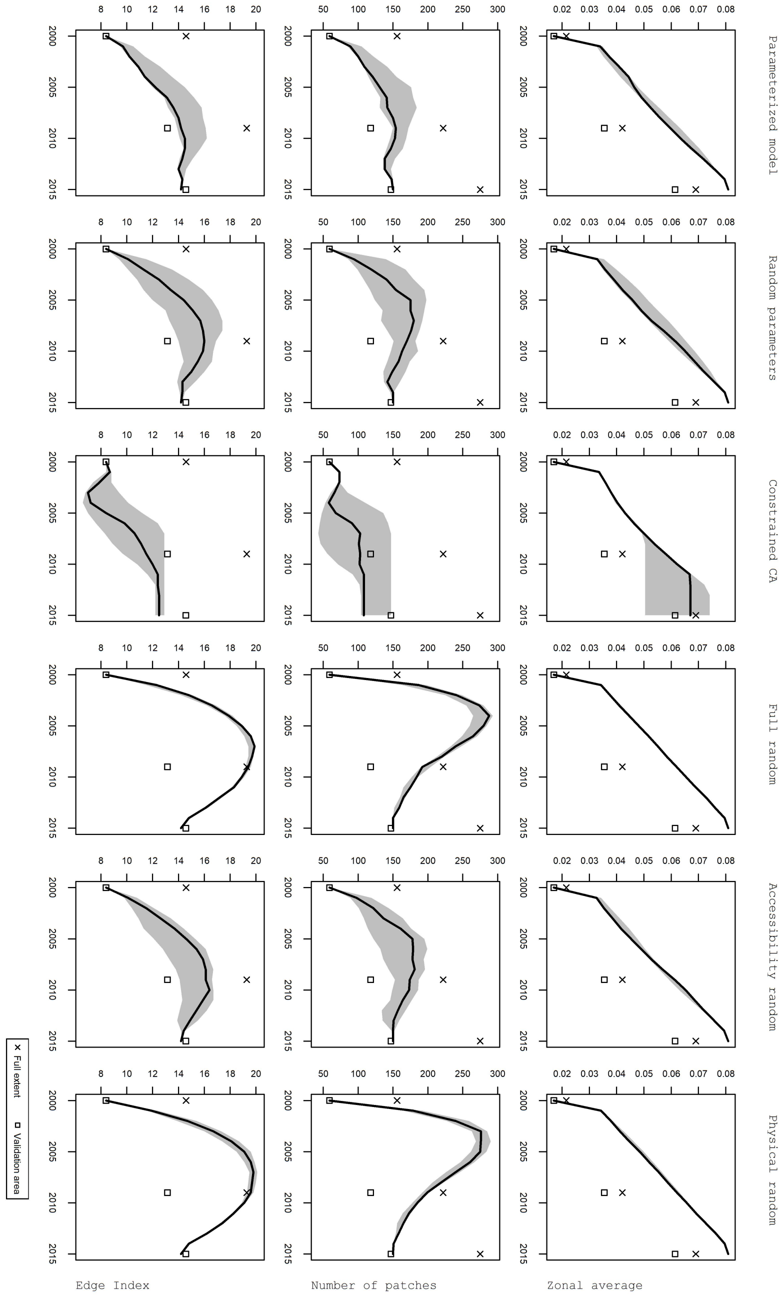


Figure 5.7 Comparison of modelled median (solid line), 95% credible interval (grey area) and observations (points) of the spatial metrics for Kigali

is the only driver, and then begins to increase.

The cause of this unsatisfactory performance of *Constrained CA* is the lack of sufficient variation in the spatial determinant of most simulated instances. The resulting posterior for the neighborhood size parameter tends to concentrate window sizes around three cells. Hence, most suitability maps for this simulation are very similar to each other and do not provide sufficient locations (cells) to fully assign land demand. The conclusion, then, is that however important the neighborhood factor is (and evidence suggests it is very important), it is not enough by itself for a good prediction. It requires ancillary information, even if limited. A line of future inquiry could be if this ancillary information must be meaningful (and the parameters with significance, as has been assumed) or if a mere random disturbance suffices and, if so, how does the model perform.

5.5 Conclusion

A method to calibrate and validate a cellular automata model of urban growth for Sub-Saharan Africa was developed. Has the method produced realistic results? It has, as the prediction of the calibrated model outperforms alternative models which randomize parameters or spatial factors. The prediction failure of models randomizing spatial factors is interpreted as evidence of a good choice of model structure, since these factors are required inputs to obtain realistic results. Additionally, the superior predictions of the calibrated model, relative to a random choice of parameters, are considered evidence of the role of calibration in improving the realism of the model outputs. One must note that the errors introduced into the model predictions by randomizing the parameters are much smaller than those introduced by randomizing spatial factors, particularly physical (i.e. not accessibility) factors.

What has been learned from interpreting the results of the parameter calibration process? The posterior distributions of model parameters – the weights of spatial factors – resulted, generally, in clear signals: it was possible to fit the derived histograms with normal distributions, most of them with reasonably sharp peaks. From the mean value of these fitted posterior distributions allowed one may establish the relative importance of spatial factors. The results support the notion that accessibility and slope have greater importance than wetlands and informal settlements location in determining urban growth.

The results of this chapter advance the use of Bayesian approaches for the calibration of cellular automata models of land, by using a Markov chain approach to determine the weights of spatial factors (ancillary information). The only preceding application of such methods to a similar modeling problem (Verstegen et al., 2014) made use of a sequential approach. While further work is required to compare both methodologies, one would expect Markov chain-based calibration to require less computational power (a distinct advantage, given the large number of iterations involved in both calibration and validation), although posterior distributions from sequential methods should present sharper signals. Regardless, the results of this chapter confirm the conclusions of Verstegen et al. (2014), that a Bayesian approach can be used to identify a cellular automata model of land dynamics; further, this chapter extends this result to a more detailed scale, a more limited spatial extent, and an urban growth context which is particularly important to current policy issues reflected in the Sustainable Development Goals.

Contrary to expectation, the calibrated model predictions (reported in the validation analysis) show a more complex landscape than that of the independently derived land cover map. This is a consequence of changes in the trend of urban growth, it is also consistent with previous evidence derived by Verstegen et al. (2016). More frequent data on land cover may be necessary to accurately characterize the evolution of land dynamics and the lack of such data, particularly when undertaking prospective simulations, represents a challenge for urban growth modeling.

The methodology here reported shows great promise for the calibration of cellular automata models. A comparison with the Particle Filter approach is important to identify when each method performs better. Finally, our cellular automata model should be tested in more case studies of Sub-Saharan Africa and the developing world more generally, particularly with different physical contexts (e.g. coastal cities), population levels, and urban extents. An important issue in these extensions is the adequacy of the selected spatial factors to explain urban growth.

5.6 Appendix

Diagnostics for Bayesian models

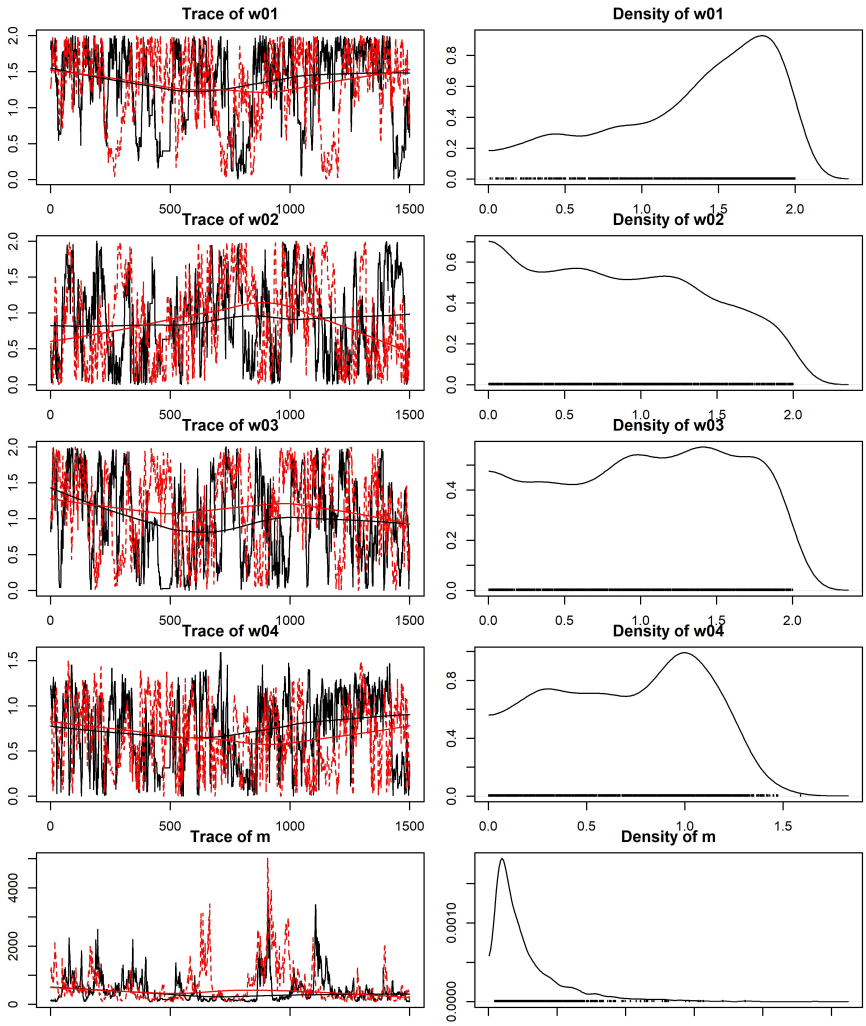


Figure 5A.1 Kampala: trace plots and density estimates

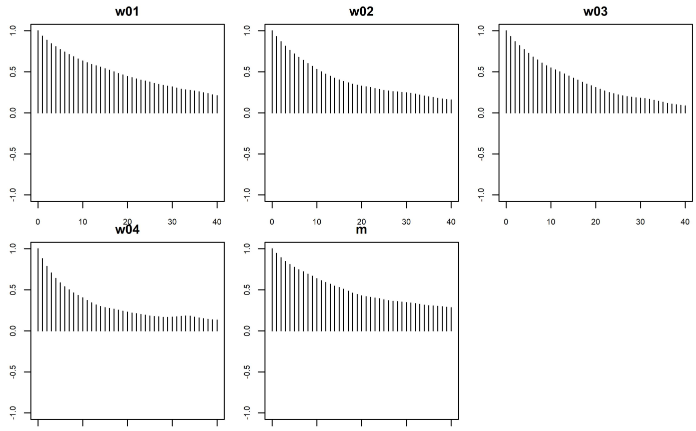


Figure 5A.2 Kampala: autocorrelation plots

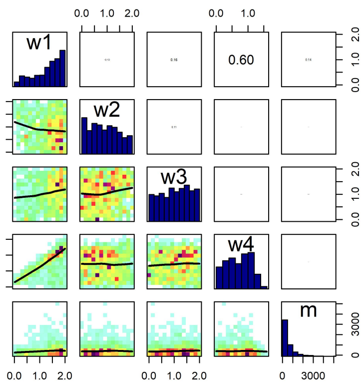


Figure 5A.3 Kampala: correlation between parameters plots

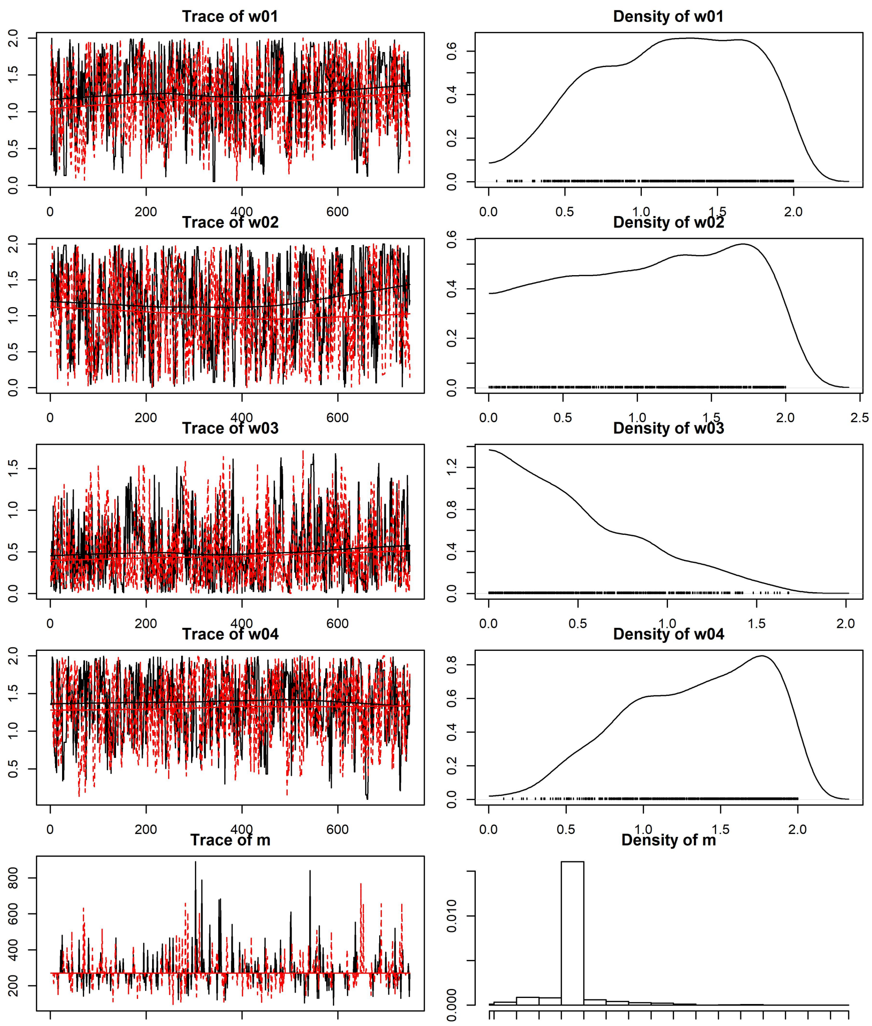


Figure 5A.4 Kigali: trace plots and density estimates

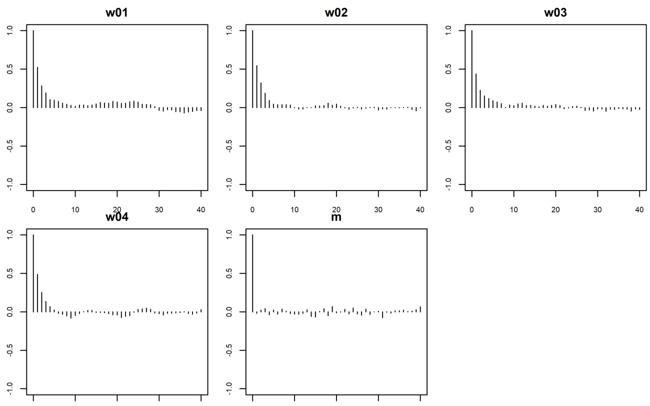


Figure 5A.5 Kigali: autocorrelation plots

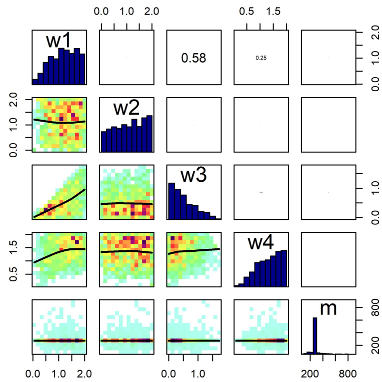


Figure 5A.6 Kigali: correlation between parameters plots

6

A causal statistical model of the impact of land regulation on urban growth

Abstract

Kigali, Rwanda is a rare case in Sub-Saharan Africa of a city that stringently applies land use controls. To what extent has this land use system contributed to shape the urban patterns of Kigali? This chapter argues it has caused a constraint on protected zones, relative to the urban zones. A difference-in-differences model was designed to assess the impact of zoning constraints on urban growth. The model was estimated using fixed effects and a spatial lag. One may conclude that, as hypothesized and despite a limited time span during which zoning constraints have been in place, the city of Kigali has managed to organize a system that significantly constraints urban development in environmentally sensitive areas. The difference-in-differences estimates of the causal effect are negative, statistically significant, and relatively large (between -0.023 and -0.030 in an area with levels of urban development between 2.7% and 4.7%). Controlling for spatial heterogeneity with fixed effects and for spatial autocorrelation – that represents the urban/non urban context in which development took place – proved necessary to obtain unbiased effects, even after including a local neighborhood effect as determinant.

Keywords: land use plan, urban growth controls, spatial causal statistical model, difference-in-differences estimator, Kigali (Rwanda)

6.1 Introduction

Despite increasing recognition of the importance of cities for the future of Sub-Saharan Africa, land use planning across the region continues to perform as poorly as in the past (Silva, 2012). In this regard, Kigali, Rwanda has been promoted as a recent positive exception of cleanliness and order (Goodfellow, 2013a). Indeed, strong evidence of strict planning implementation exists (Goodfellow, 2013b), although critical perspectives have argued these plans were developed without taking into account Kigali's context and, thus, they do not respond to the city's needs (Watson, 2014). Regardless of judgment on how appropriate the regulation may be, it is clear it should have caused a measurable effect on urban patterns (due to its strict implementation); in this sense, Kigali is a rare opportunity to explore the actual impact of regulation on the physical characteristics of a city in Sub-Saharan Africa, and generally in the developing world.

This chapter is concerned with the impact of Kigali's land use regulations on urban growth patterns. A series of policy and law instruments were implemented in Kigali after 2000 (Goodfellow, 2012), which constrain urban development from environmentally sensitive areas and purportedly promote densification of more suitable for land. Development decisions are measured (as originally proposed by Dempsey and Plantinga, 2013) with mid-scale land cover maps that identify built-up and non built locations, using the Global Human Settlement Layer (GHSL) of the European Union's Joint Research Center (Pesaresi et al., 2016); this detailed spatial scale enables greater precision in quantifying effects relative to aggregate data. Data organized into a panel structure allowed for the estimation of a difference-in-differences model to determine the causal effect of the regulation; the model was constructed using a spatially explicit causal framework proposed by Kolak (2017) to control for and understand any spatial autocorrelation present.

6.2 Land use regulation effects on urban patterns

6.2.1 Land use planning in Kigali

When looking at the evolution of Kigali's urban morphology, by 2000 urban patterns had been configured in the relatively short time span since the 1994 genocide, mostly through slums to house hundreds of thousands of Rwandans: a first wave of Tutsi refugees and a second, the returning inhabitants of Kigali who had fled in the aftermath of the genocide (Goodfellow, 2012). During the following decade, a series of urban policy instruments was introduced in Rwanda and Kigali (Goodfellow, 2012). As part of this effort, the Kigali Conceptual Master Plan (2007), the Building Control Regulations (2008) (Goodfellow, 2013a), and the Kigali Master Plan (2013) were produced and implemented. Given how strictly these regulations have been applied and how fast the city has expanded, it is likely urban patterns by 2014 had been at least partially determined by the land use regulations.

There is a conceptual continuity between actions and policy instruments in Kigali over this period. By 2006, the city was already restricting development in wetland areas (Goodfellow, 2013a). The Conceptual Master Plan (OZ Architecture et al., 2007) is a strategic plan containing a vision of Kigali that realizes the principles of compact growth, namely constraints to urban development in the periphery and encouragement of densification in central locations. These ideas, which had associated locations in the Conceptual Master Plan, were refined in the Kigali Master Plan: zone boundaries were defined at more detailed spatial scales and building regulations were parameterized through floor-to-area ratios (FAR), maximum building coverage, and other design regulations.

Kigali's land use regulations have been criticized for responding more to the interest of real estate investors than to the city's residents (Watson, 2014). The urban poor have been displaced, in general suffering impoverishment (Nikuze et al., 2019), to make way for high rise re-develpoment projects. Additionally, serious doubts exist about the feasibility of developing such high rise buildings as the regulation envisions (Pérez-Molina et al., 2016; Behuria and Goodfellow, 2019). However, there is also strong evidence that natural systems have been protected and urban regulations have been strongly enforced (Goodfellow, 2013a). What impact should be expected of these conditions?

6.2.2 Planning in the configuration of land use patterns

When analyzing the effect of regulation, this chapter develops the view that a constraint on urban development is the main relevant phenomenon in Kigali. In this sense, one can interpret the dichotomy between urban zones and the protection zones in the Kigali Master Plan as an urban growth control. Urban growth controls (see Anthony, 2017 for a description of such strategies) are policies designed to ameliorate market failure arising from externalities in land and housing markets, as urban development does not internalize all the costs to society of expanding the urban fabric of a city. New infrastructure that makes development possible – and its use by new urban land users – is generally subsidized, for example, nor are the opportunity costs of environmental and agricultural systems fully accounted for (Brueckner, 2000).

Since urban growth management policies seek to internalize these costs, they consequently lead to higher housing costs and to less development. Anthony (2017) proposes five such effects (arguing fundamentally from the US perspective), of which two are relevant for Kigali: a land-scarcity effect because much land is designated as protected (from development) and an amenity effect in the protected area. The first constrains urban development in the controlled area but the second promotes it, since areas without much urban development have advantages in terms of the quality of their environment and the absence of negative externalities from congestion.

The effect of such urban management policies, then, is less straightforward than one may anticipate and may have unintended consequences on multiple dimensions. Different policy instruments may cause contradictory outcomes and, furthermore, methodological decisions may also substantially change the estimates of these effects (e.g. see Jackson, 2016, who analyzed critically these issues for the state of California, US). For example, whereas Jackson (2016) identifies zoning as the most efficient constraint on development, Glaeser and Ward (2009), analyzing the city of Boston, identify minimum lot requirements as the main policy vehicle to restrict supply of urban land there – leading to higher housing prices and less construction, with most development occurring in the denser areas. They conclude the city has adopted land use controls that are suboptimally restrictive (Glaeser and Ward, 2009). Kahn et al. (2010) and Severen and Plantinga (2018) use regression discontinuity and difference-in-differences estimators (causal statistical models) to calculate a 20% increase in single-housing unit prices within the California coastal boundary zone, which are subject to constraints that improve amenities and reduce land supply, although they also impose additional costs on property owners. Dempsey and Plantinga (2013) analyze urban growth boundaries in the state of Oregon's Willamette Valley, US, with a difference-in-differences estimator and conclude they did in fact restrict development, although not to the point of limiting population growth.

Other regulatory constraints, when not appropriately designed, may also have detrimental effects on urban land patterns. As in Glaeser and Ward (2009), Brueckner and Sridhar (2012) argue Indian cities have excessively constrained urban development – in this case, by promoting low-rise land use patterns through low FAR. The consequence has been an excessive horizontal expansion of cities, leading to social costs in the form of increased commuting costs (relative to a higher city counterfactual). A similar point was raised by Cai et al. (2017) for Chinese cities, although in their case, developers leveraged political power to exceed the restrictive FAR, rendering the regulation ineffective.

In view of this discussion, what may one expect for Kigali? Fundamentally, a constraining effect: FAR in Kigali allows for excessively high buildings, any constraints on high-rise development are likely technological and related to lack of demand. On the other hand, a strict implementation of urban development controls should reduce urban growth in protected zones and, overall, cause greater compaction of the city's urban footprint.

6.3 Methodology: spatial difference-in-differences estimator

6.3.1 Statistical modeling of causality: difference-in-differences technique

A difference-in-differences estimator is proposed to quantify the effect of zoning constraints on urban growth in Kigali. The difference-in-differences estimator is part of a long tradition in the econometric evaluation of policy (Imbens and Wooldridge, 2009), going back to the seminal work of Ashenfelter and Card (1985).

Baum-Snow and Ferreira (2015) conducted a recent review of applications of causal statistical models in urban and regional planning; they found a wide variety of applications (to labor, housing, infrastructure variables, among others) that allowed them to conclude any application requires many specific methodological decisions, subject to the research question and available data. However, they do argue two questions are central to causal analysis: to understand the sources of variation in the treatment variable and to recognize which treatment effect is being estimated, given several possible methodological approaches to explore the same question.

In this vein, the main methodological antecedent of this chapter is the work of Dempsey and Plantinga (2013), who use a spatial statistical model (a model for which the records in the data set are locations, in the case of Dempsey and Plantinga, 2013, plots) for an outcome variable of development/no development and an urban growth boundary (a zoning constraint) as the treatment variable. Dempsey and Plantinga (2013) use econometric techniques robust to heteroskedasticity but avoid explicit spatial correlation because they do not know the true form spatial interdependence in the data. Consequently, they argue spatial correlation poses too great a risk of mis-specification.

The difference-in-differences estimator was implemented following the framework of Kolak (2017), who proposed the design of the model should consider explicitly what spatial effects are present and how they must be formalized (see table 6.1 in section 6.4 for the application of the framework). Difference-in-differences is an estimator that implements the counterfactual framework of causal analysis, which relies on the Stable-Unit-Treatment-Value-Assumption: that outcomes are independent of actual treatment assignment; this assumption is very likely to be violated by spatial data – because spatial data represent phenomena located in time and space that are subject to spatial interdependence – and, because of this, difference-in-differences estimators applied to spatial data can be very easily biased. Kolak (2017) proposed a line of argument to understand the research problem as a way to incorporate spatial effects where and when needed to control for this type of bias.

With this in view, a database of 2125 locations was compiled for two years (2000 and 2014): urban development was taken as present or not and defined as the dependent variable; from the Kigali Master Plan, in turn, areas constrained from development (protection zones) and areas acceptable for urban development (residential, commercial, industrial zones) were derived. The data, thus, was structured as a two-period balanced panel.

This database, which is fully described in subsection 6.3.2, was employed to estimate a fixed effects model – the fixed effects capturing the spatial heterogeneity in the outcome variable – with instrumental variables, because the treatment (zoning constraints) is thought to be endogenous:

$$Y_i = B \cdot D_i + \beta_i \cdot X_i + \epsilon_i \tag{6.1}$$

$$D_i = \gamma_1 \cdot Z_i^1 + \gamma_2 \cdot X_i + \omega_i \tag{6.2}$$

with Y_i the outcome variable (urban development), D_i the treatment variable (equal to 1 in constrained zones and period 2014, and to 0 for all other records), X_i a control (the neighborhood effect, average urban development within a moving window of 7 cells; this variable was transformed to its natural logarithm, adding 0.01 to each value to avoid null data from 0.00 values), and Z_i^1 a dummy variable equal to 1 for wetland areas and to 0 otherwise. The causal treatment effect is represented by B.

The sensitivity of the model to different estimation methods (fixed effects vs. random effects, the presence or absence of a spatial lag) were explored and are reported in the results.

Equation 6.1 was estimated using the the Generalized Moments tool *spgm* of the *splm* package (Millo and Piras, 2012) of R (R Core Team, 2017). When testing sensitivity to methodology, the *gmm* package (Chaussé, 2010) was used to estimate the pooled model with no spatial effects and the *spml* tool from the *splm* package was used for the pooled version with spatial lag (with the treatment variable was predicted using a OLS linear model).

6.3.2 Study area and data models of spatial patterns

The modeled extent of Kigali corresponds to the area simulated in chapter 5. When overlaying the regulated area, the administrative limit of the Kigali province, the southeast and northeast corners of this extent were found to lie outside of the province and were, in consequence, excluded.

To build the final sample, first a uniform sample for each year was constructed by taking one location out of every five columns or rows; this should reduce spatial autocorrelation in the data (although it is unlikely to eliminate spatial autocorrelation completely). The uniform grid of sampled locations consisted of 24426 records for each year. Second, all locations corresponding to roads or infrastructure (planned and existing) were excluded, since they do not represent social dynamics of urban actors. The zoning maps of both 2025 and 2040 include these areas. After filtering out these locations, the database included 21531 records per year. A random sample of approximately 10% of these was selected and the data for both periods was pooled, resulting in a sample of 4250 records structured as two repeated cross sections, one for 2014 and the other for 2000.

For each one of these locations, the following variables were sampled:

Land cover: land development decisions for 2000 and 2014 were obtained from the GHSL (Pesaresi et al., 2016), a Joint Research Center contribution that produces world-wide maps of built-up areas at a resolution of 38m to monitor the location and size of human settlements with applications to quality of life (accessibility), exposure to hazards and pollution, and environmental impact. Human settlements data was resampled to 30m, since it corresponds to the inputs of the land cover models described in chapter 2. Urban development, the outcome variable of the difference-in-differences model, is equal to 1.00 for locations where the GHSL identifies buildings and 0.00 for other locations.

Zone: a dichotomic variable equal to 1.00 for the protected zones (agricultural land, parks and recreation, environmentally sensitive areas) and to 0.00 for the urban land uses (commercial, residential, and industrial zones) of the Kigali Master Plan (SURBANA International Consultants PTE Ltd., 2013), which coincides, as noted, with the Kigali Conceptual Master Plan of 2007 (OZ Architecture et al., 2007). The $Zone$ variable is the treatment D_i in the difference-in-differences model. The 2040 zoning map of the Kigali Master Plan was used to build this variable. Note that while development is not completely forbidden in the protected zones, it is highly discouraged: only support buildings (e.g. barns, recreational facilities) are allowed with total building areas strictly limited.

Neighborhood factor: the average built-up land cover in a moving window of seven cells (210m), estimated using the land cover data as input. The variable controls for the differences in urban growth between urbanized areas and non-urbanized areas, which are important in Kigali because infrastructure constraints and a relatively sparse urban fabric make infill development attractive. Furthermore, this neighborhood factor controls for scaling relations (Batty, 2009) embedded in variables such as accessibility. The natural logarithm of the neighborhood factor was used as control in the regression models.

Wetland: a dichotomic variable equal to 1.00 for wetland and river areas, identified in the zoning map, and to 0.00 for all other locations; while these areas should in principle be unsuitable for urban development, what is termed wetland often corresponds to the generally dry floodplains of very small rivers. The physical characteristics of the landscape result in buildings being feasible, if undesirable for flood risk and other environmental reasons. Additionally, this variable is meant as an instrument for the $Zone$ treatment variable – and since it was one of the criteria used to determine protected areas, it is thought to be an good instrument.

Other variables had been identified in chapter 5 as relevant to explain urban growth: the slope factor, travel time to CBD, and distance to main roads. However, these variables have hardly any variation in time – and the data models which represent them in previous models are also static in time. Because this chapter reports on the application of fixed effects models, such variables would be generally dropped from the analysis.

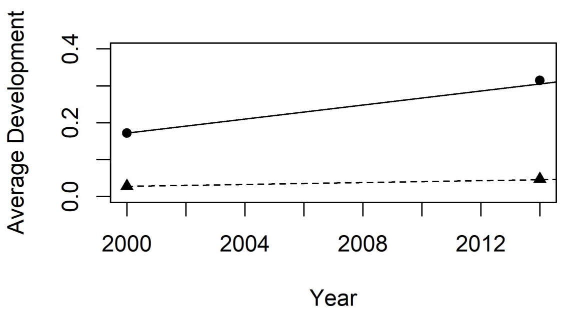


Figure 6.1 Change in average development over time. Dotted line corresponds to *Zone* (constrained area)

6.4 Results and discussion: difference-in-differences model

A preliminary aggregate characterization of urban development in Kigali suggests positive impacts of land use planning on urban patterns. Figure 6.1 shows the average development for the constrained area (dotted line) and the non-constrained area.

Two important trends are shown in figure 6.1: (1) development in Kigali increased between 2000 and 2014 in both the constrained and the non-constrained area, but it increased more in the non-constrained area; (2) the amount of development is much greater in the non-constrained area than in the constrained area. On aggregate, this description suggests land use regulations have been effective in shaping urban patterns towards a more compact, sustainable arrangement. However, this descriptive analysis does not account for space, in particular spatial heterogeneity. The non-constrained area includes urban centralities and the area most accessible to them; thus, the results of figure 6.1 confound spatial dependence patterns with the potential effect of the policy.

The difference-in-differences estimates are presented in tables 6.2 (which explores sensitivity to choice of methodology) and 6.3 (which tests the sensitivity of models to the spatial lag). Table 6.1 presents the application of Kolak's (2017) spatial framework.

Three key methodological choices follow from table 6.1. First, because the zoning scheme was defined using the same criteria that determine urban agents' location choice, the treatment is not exogenous to the outcome. This condition implied the need to use instrumental variables and, therefore, a General Moments estimation method. Second, fixed effects were chosen to control for spatial heterogeneity. The most relevant spatial differentials embody the scaling relations that describe urban morphology: a neighborhood effect and accessibility to urban centralities.

Since estimates of accessibility are all approximately constant in time (and therefore would be dropped by fixed effects estimators), only the neighborhood effect was controlled for. Third, spatial lags were not expected to play a substantive role within the model, since the zoning treatment was externally imposed and it does not change in time nor with location. All records within $8000m$ of a location were defined as its neighbors; this limit was chosen by fitting the urban growth data of the sample to a semivariogram and establishing its range.

The treatment variable *Zone* of the models reported in tables 6.2 and 6.3 was instrumented using the *Neighborhood factor* and the *Wetland* variables. As noted previously (Pérez-Molina et al., 2019b), six spatial determinants explain urban growth in Kigali: neighborhood effects (from these results, occurring within a moving window with sides of $220m$), travel time to CBD, Euclidean distance to main roads, slope, and wetlands. Of these, only the neighborhood factor changes in time. Furthermore, for theoretical (Glaeser, 2008) and empirical reasons (Pérez-Molina et al., 2019b), one should argue accessibility factors are the most important; they are also highly correlated with neighborhood effects.

The difference-in-differences estimators are shown in table 6.2, they correspond to the regression coefficients of the variable *Zone*: all models result in a significant and negative causal effect, which means the constrained areas present less urban development than non-constrained in Kigali. However, the pooled regressions – those that do not control for spatial heterogeneity – result in a much larger causal effect than the fixed and random effects models. This result is understandable, since (despite

Table 6.1 Application of spatial framework

How is treatment chosen or assigned?	Treatment is exogenously imposed, top-down, on locations; the criteria to define the treatment area make this variable endogenous to the response (urban planner judges the most suitable locations for urban development by 'replicating' the choices of rational urban agents).
What are potential sources of variation in the treatment variables?	Sources of variation are systematic and suitability-informed; in this sense, policy is the consequence of spatial heterogeneity. No spatial lag expected, though, since the policy is imposed top-down.
What effects are being estimated (if any)?	Causal effect of treament (Zone) Fixed effects (group and temporal) Instrumental variables (adds wetland location to accessibility controls of model)

Table 6.2 Regression estimates for panel data analysis to explain urban growth. Sensitivity to estimation methods

Variable	Fixed		Random		Pooled	
	No Lag	Sp. Lag	No Lag	Sp. Lag	No Lag	Sp. Lag
Intercept	–	–	0.712	0.691	0.736	0.736
			(< 0.01)	(< 0.01)	(< 0.01)	(< 0.01)
Zone	−0.014	−0.041	−0.017	−0.020	−0.129	−0.129
	(0.029)	(< 0.01)	(0.004)	(< 0.01)	(0.052)	(< 0.01)
Neighb.	0.140	0.124	0.158	0.155	0.155	0.155
Effect	(< 0.01)	(< 0.01)	(< 0.01)	(< 0.01)	(< 0.01)	(< 0.01)
Spatial Lag	–	0.478	–	0.074	–	−0.0001
		(< 0.01)		(0.044)		(0.362)
Hausmann test						
No spatial effects	$\chi^2 = 1436.3$, 1 d.f. (< 0.01)				Reject RE.	

All estimates using GM estimator except for *Pooled Spatial Lag* model, which uses a maximum likelihood estimator.

the use of instrumental variables to address the endogeneity problem of the variable *Zone*) the fixed and random effects are the principal method to control for spatial heterogeneity. When such controls are not included, the *Zone* variable does not represent only the causal effect of the land use plan; it also serves as a proxy variable for accessibility, because the areas constrained from development are farther from urban centralities than the non-constrained areas.

Apart from the causal effect, the other major difference between models is in the spatial lags: it is not significant for the pooled model, and positive and significant for both the fixed effects and random effects models but the spatial lag of the fixed effects model is over six times larger than the random effects. Other regression coefficients are all significant, positive, and of similar magnitude for all models (the neighborhood effect for all six versions of the model and the intercept for the random effects and pooled models). Furthermore, the causal effect is similar for the fixed effects and random effects models, less than −0.10 in all cases.

Finally, the Hausmann test to contrast the fixed effects and the random effects model rejects the consistency of the random effects model. Thus, one should favor the fixed effects.

While the model design assumes spatial lags should not be part of the model, it is important to test the robustness of the model to this assumption (recall the caution of Dempsey and Plantinga, 2013 regarding spatial autocorrelation). Table 6.3 shows how these spatial lags affect the difference-in-differences estimator: as the radius that determines the spatial lag decreases, so does the magnitude of the causal effect, although all causal effects are negative and statistically significant. The spatial lag

Table 6.3 Regression estimates for panel data analysis to explain urban growth. Standardized coefficients reported. Sensitivity to estimation methods

Variable	Spatial Lag				
	No Lag	8000m	4000m	2000m	1000m
Zone	−0.014	−0.041	−0.030	−0.025	−0.023
	(0.029)	(< 0.01)	(< 0.01)	(< 0.01)	(< 0.01)
Neighb.	0.140	0.124	0.121	0.119	0.113
Effect	(< 0.01)	(< 0.01)	(< 0.01)	(< 0.01)	(< 0.01)
Spatial Lag	–	0.478	0.421	0.378	0.373
		(< 0.01)	(< 0.01)	(< 0.01)	(< 0.01)

Fixed effects models with Spatial Lag. Spatial Lag defined as all records within a radius of 1000 to 8000m of each location. All estimates using GM estimator.

itself, also statistically significant and positive for all models, decreases with the radius as well, and the neighborhood effect is approximately the same for all models.

Methodologically, there are two important issues that must be raised from the results of this chapter: the choice of linear regression methods, despite the dependent variable being dichotomic, and the relevance of spatial autocorrelation.

The choice of linear over non-linear econometrics, also adopted by Kline et al. (2014), was explained by Dempsey and Plantinga (2013): non-linear econometrics (probit or logit models) have the advantage of constraining the dependent variable to the range $[0, 1]$; however, the causal effect estimated is a function of all variables and parameters and it becomes biased in the presence of heteroskedasticity. Ultimately, since the center of this chapter's results is analytic rather than predictive (interest focuses on the causal effect rather than the predictions of the model), the linear estimator was chosen.

Regarding the spatial lag, there is a potential contradiction between the results reported in table 6.3 and the assumptions made when designing the difference-in-differences model. Should a spatial lag be included in the model? What radius should determine it? As was noted, the spatial lag of 8000m is the result of an empirical analysis (of a variogram); does this radius have a substantive interpretation? A spatial lag can be generally interpreted as evidence of a diffusive process. One should note the neighborhood factor already controls for such a process at a very local scale (seven cells, 210m).

To address this question, figure 6.2 shows the urban development patterns and the radius that determine the spatial lag for three locations: one in the main urban patch, one town in the eastern edge of the study area, and one in undeveloped regions. What one can see in these four circles is that the three smallest (for which the radius varies from 1000m to 4000m) show a similar pattern within them: for the urban patch, an urban context, and for the locations in the eastern town and northern

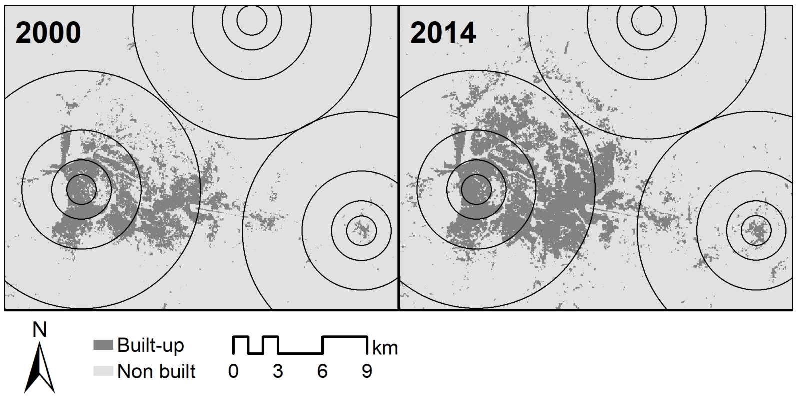


Figure 6.2 Urban development patterns and radius determining spatial lags for four locations, 2000 and 2014

undeveloped location, the essentially undeveloped hinterland of the city. Therefore, it would seem that these spatial lags control for the mid-scale context of each location (whether they are in the city proper or in the surrounding periphery). The band of area between the third and fourth circles (radius of $4000m$ and $8000m$) is different. In all three cases, the overall proportion of area in this band differs from the inner area: for the urban location, the largest area is undeveloped (contrary to the inner circle), and for the peripheral locations, while the undeveloped area is still dominant, relatively large patches of urban development can be seen.

Therefore, the positive spatial lag – contrary to the initial assumption of the methodological design – is required to control for the general context. In addition, the spatial lag is defined by all locations within a radius of $1000m$ to $4000m$. When looking at table 6.3, the causal effect of the zoning is a net reduction of urban development between 2.3% and 3.0%, relative to non constrained locations. This is, contrary to initial appearance, an important reduction, given the average build-up area in the constrained area was a mere 2.7% in 2000 and 4.7% in 2014.

When compared to previous cases reported in the literature with similar methods (i.e. a difference-in-differences estimator drawn from a spatial statistical model), as in Dempsey and Plantinga (2013) or Kline et al. (2014), the causal effect of regulation in Kigali seems small (Dempsey and Plantinga, 2013 estimated net increases of urban development within the urban growth boundary of 12.7%; Kline et al., 2014, of 18.7%). However, this conclusion should be qualified: first, as noted, because the percentage of development in the area constrained by zoning is also small and second, neither Dempsey and Plantinga (2013) nor Kline et al. (2014) use fixed effects; their results are closer to the pooled model reported in table 6.2 – for which the (biased) causal effect is a reduction of development of 12.9%; interestingly, they do control for spatial heterogeneity but by introducing exogenous spatial determinants rather than spatial econometrics.

A second difference between these two case studies and the evaluation of Kigali lies in the time extent involved. Both Dempsey and Plantinga (2013) and Kline et al. (2014) evaluate a long-standing planning instrument (urban growth boundaries enacted in Oregon and Washington, US, in the 1970s) over a much longer period (1970s to 2000s) and in a more mature institutional setting. By comparison, the analysis of Kigali reported in this chapter covers a period of 15 years and the regulation was in place for half of the period. It is also true, however, that in Kigali, the total population and urban footprint doubled during this period; a very large change occurred compressed in time, even after the upheaval associated with the 1994 genocide had subsided. In this sense, several cities for which Dempsey and Plantinga (2013) reported causal effects also presented similar population growth rates but for longer time periods.

A third element to consider is the very wide heterogeneity of results obtained in previous studies (Dempsey and Plantinga, 2013; Kline et al., 2014). Of 17 Oregon cities analyzed in Dempsey and Plantinga (2013), 5 resulted in non-significant causal effects of their urban growth boundary, 6 presented a small constraint (less than 10%), and 6 a large constraint (over 10%). Even in Kline et al. (2014), who analyzed a single metropolitan area, one of four counties showed a non-significant effect. These results, as well as those reported in this chapter (specifically the difference between the pooled model and the models controlling for fixed or random effects in table 6.2) strongly suggest the presence of important spatial effects that, even after controlling for main determinants such as accessibility, may introduce idyosincratic elements with the potential to bias results.

6.5 Conclusions

This chapter has examined the extent to which zoning constraints embodied in the Kigali Conceptual Master Plan and the Kigali Master Plan have, over the period 2000 to 2014, restricted urban development in the city of Kigali. While most urban development over this period occurred within areas designated for urban use in the planning instruments, this can be partially attributed to spatial heterogeneity (specifically, to the fact that urban zones in the plans are more accessible than environmentally sensitive areas in the periphery). However, using a difference-in-differences model that explicitly accounts for spatial effects, the net effect of zoning was estimated to be statistically significant and to range between -2.3 and -3.0 (with average built-up fraction increase in the constrained area from 0.027 to 0.047).

Methodologically, a fixed effects model with instrumental variables and spatial lag was designed, following the framework proposed by Kolak (2017). Instrumental variables were deemed necessary because accessibility, the main theoretical determinant of urban development, is clearly correlated with and conceptually contributes to determine the zoning scheme proposed in the land use regulation. Spatial lags were initially thought to be unnecessary, as local diffusion processes had already been controlled for using a neighborhood factor. However, upon further critical examination, an effect of mid-scale level context (whether urban or rural) was found to be captured by spatial lags with neighbors determined by ranges between $1000m$ and $4000m$. Explicit consideration of spatial effects, both methodological and substantive, was found to be necessary to avoid biased estimated in the case of Kigali.

Kigali is a rare case of a city in a developing country with a strong institutional setting. It is an open question whether the social and political context that has made such strict application of regulations will persist in time (Goodfellow, 2013b). Furthermore, if urban development continues apace and it is not matched by economic prosperity, the envisioned supply of urban space will not materialize, causing a potential breakdown of the system – and this is a clear danger in the short run. Paradoxically, looking at the next decade, policy makers may want to incorporate flexibility into their planning instruments to cope with such contingencies. Their challenge may very well be to incorporate such flexibility without renouncing the environmental and infrastructure provision advantages that follow from compact development.

6.6 Appendix

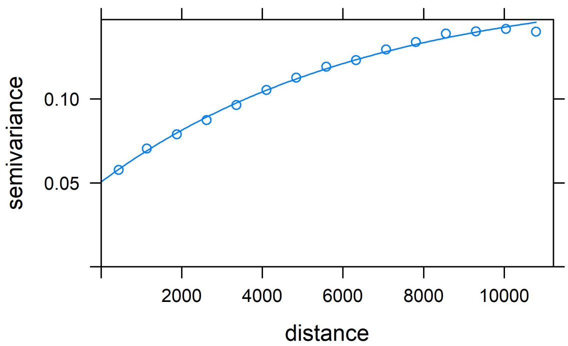


Figure 6A.1 Variogram of urban development of final sample used to determine spatial lag. Exponential fit with range of $8000m$ and psill of 0.15 and nugget of 0.05

7

Flooding, land cover, and spatial planning scenarios for Kampala and Kigali

Abstract

Prospective scenarios of urban growth and flooding were designed and developed to explore the relations between land use regulation, urban growth, and flood impacts. Critical uncertainties determined the design of scenarios: land demand (how much population growth may occur in each city and can it be physically accommodated), potential supply (will growth follow planned or unplanned conditions, and if unplanned, under what densification preferences of urban agents), and rainfall (which rainfall events will influence urban patterns). The calibrated cellular automata urban growth model of Kampala and Kigali was used to project urban growth to 2030. Scenarios assuming no feedback effects between flooding and urban growth were estimated to understand the potential role of land use regulation of Kampala and Kigali, in terms of the resulting urban pattern and of its potential flood impacts. A further set of scenarios incorporating a feedback effect between flooding and urban growth, under planned and unplanned conditions, were developed for Kampala. Results of these scenarios were: for Kampala, land use regulation promotes compact growth but the land use plan of Kigali may have the unintended consequence of promoting sprawling patterns, land use planning in Kampala may significantly reduce exposure to flooding but not runoff, incorporating the feedback effect visibilizes the benefits of land use regulation for large cities (but this reduction in flooding is unlikely to be present in smaller cities, such as Kigali).

Keywords: land scenarios, flooding, urban growth, land use planning, flood risk mitigation, Kampala (Uganda), Kigali (Rwanda)

7.1 Introduction

The scenarios developed in this chapter for Kampala and Kigali revolve around the research questions which organize this dissertation: (1) the interaction between urban growth and flood processes and (2) how can these be influenced through land use planning.

Loosely following the method proposed by Shearer et al. (2009, chapter 6), around this focal question, *land use planning* was deemed to be the main local force that could best contribute to urban flood risk mitigation and *land demand*, *urban land supply*, and *rainfall* were identified as the main contextual uncertainties. Therefore, the logic of the scenarios concentrates on exploring the impact of these elements: in a first stage, different population projections and urban development characteristics are simulated from 2016 to 2030 using planned and non-planned conditions (for Kampala and Kigali, see section 7.3); in a second, the feedback between flooding and urban growth was introduced into the land use modeling, with flooding a consequence of rainfall trends (only for Kampala, see section 7.4).

Variations of the urban growth model reported in chapter 5 were used to construct the projections, which were then evaluated and discussed in terms of the resulting urban development patterns, flood outcomes, and specifically how these differ between the planned and unplanned scenarios. Overall, prospective scenarios not only organized data to explore the relations, they also visibilized assumptions affecting these relations, setting explicit conditions against which to test the plausibility and possibility of each simulation.

7.2 Critical uncertainties

7.2.1 Land demand

The determination of land demand for prospective scenarios follows the same procedure as was outlined in the prototype model of chapter 4, namely by projecting an exogenous population growth and dividing it by gross population density. Under the most basic assumption, gross population density for any target year was taken as equal to the baseline year (2015 for Kigali, 2016 for Kampala): 128.7 residents per ha for Kampala and 251.6 residents per ha for Kigali. Note this estimates use the land cover models developed in chapter 2, which are of built-up fraction and thus result in estimates of urban area that are lower than other studies of Kampala or Kigali.

The amount of future population in Kampala and Kigali is an uncertain parameter. Eastern Sub-Saharan Africa is generally considered a region undergoing very rapid urbanization (Dodman et al., 2017), often characterized as *catching up* to other parts of the world as the fraction of urban population increases; however, evidence of slowing urbanization rates in Sub-Saharan Africa had already been advanced ten years ago,

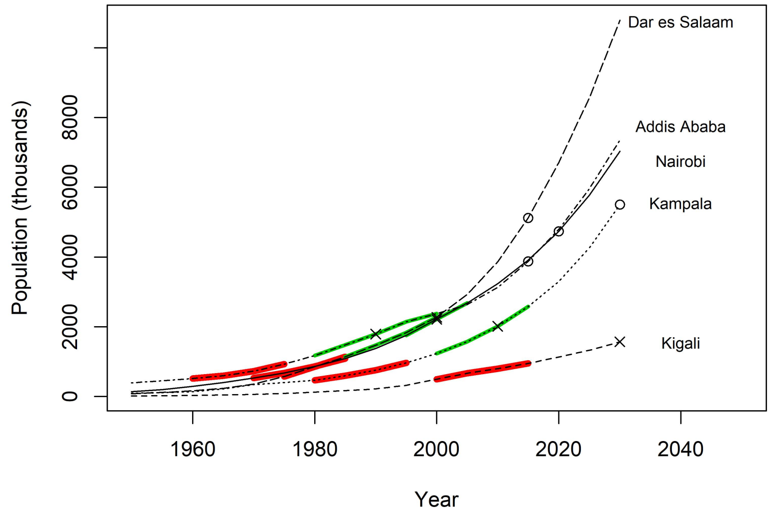


Figure 7.1 UN population estimates and projections of Eastern Sub-Saharan Africa cities (United Nations, 2018). In red: period of growth between circa 500 thousand and 1.0 million residents; in green, period of growth between circa 1.0 million and 2.5 million residents; crosses: population 15 years after reaching circa 1.0 million residents; circles: population 15 years after reaching 2.5 million residents

heralding changes in the determinants of regional urban growth (Potts, 2009). The underlying argument, for the future, is that urban population – and consequently the population of major cities – will continue to rapidly increase at similar rates to the immediate past. This can be seen in figure 7.1, which plots population estimates and projections for five year intervals, compiled by the United Nations (United Nations, 2018), of five major cities of Eastern Sub-Saharan Africa: Addis Ababa (Ethiopia), Dar es Salaam (Tanzania), Kampala (Uganda), Kigali (Rwanda), and Nairobi (Kenya).

A more careful analysis of the data on figure 7.1, however, reveals at least two trends in the data: while Dar es Salaam, Nairobi, and Addis Ababa since 2000 exhibit the exponential growth characteristic of Kampala, Kigali and Addis Ababa before 2010 expanded at a slower pace. A deep analysis of the causes of these demographic differences is beyond the scope of this chapter (refer to United Nations, 2018 and Dodman et al., 2017 for general context as well as the arguments of Potts, 2009). However, it does suggest at least two hypothesis of population growth are necessary: a high population growth and a low population growth:

- For **Kampala**, trend urban growth corresponds to a rapid increase with interannual rates over 5.0%. The high population growth was defined as the trend condition, which predicts a population of 5.5 million residents by 2030.

- For **Kigali**, trend urban growth corresponds to a "slow" (around 3.0% yearly) increase. The trend condition was taken as the low population growth – by 2030, a predicted population of 1.6 million residents.

What should be the low population projection for Kampala, the high population for Kigali? In 2000, Kampala's population was estimated at 1.2 million residents and Kigali's, at 500 thousand residents. Over the next 15 years, these increased to 2.6 million and 951 thousand residents respectively. Figure 7.1 shows in green the increase from circa 1.0 million to circa 2.5 million residents of Addis Ababa, Dar es Salaam, and Nairobi; the circle marks the period 15 years after these expansions took (Dar es Salaam, Addis Ababa) or are projected to take place (Nairobi, Kampala). As is shown, Dar es Salaam and Nairobi result in estimates or projections that are approximately equal to Kampala's projected 5.5 million. However, Addis Ababa's population was substantially lower, at 3.8 million residents. Therefore, a low population growth condition for **Kampala** was assumed as a projected population equal to that of Addis Ababa in 2015. Similarly, the red lines of figure 7.1 mark the increase from circa 500 thousand to circa 1.0 million residents for the five cities that have been discussed and the cross, the estimate or projection 15 years after the 1.0 million residents mark has been reached. As in the previous case, Addis Ababa and Kigali show a lower and similar estimated or projected population (around 1.5 million) while the faster growing cities (Dar es Salaam, Kampala, Nairobi) also show similar and larger population estimates of over 2.0 million residents. Thus, the population of Nairobi 15 years after passing 1.0 million residents (2.2 million residents) was adopted as a high population growth estimate for **Kigali**. Exact figures are reported in the appendix, with other scenario parameters.

An assumption in the discussion on projected population is the capacity of the environment of accommodating the land demand under trend conditions, which in turn is grounded implicitly on the *closed city assumption*, namely migration cannot occur (by which one should understand that an urban agent cannot leave the city if it becomes too crowded; see Brueckner, 1987 for a formal neoclassical analysis of urban structures under closed and open city assumptions). What could happen if such development is not physically possible? At least two possible system responses are predicted by the theory (Brueckner, 1987): under closed city conditions, the city would densify (its capital-to-land ratio would increase) and this densification would generally be greater in more accessible locations. Alternatively, under an open city assumption, the increased costs of congestion (the negative externalities of densification) would reduce utility, in turn prompting emigration; thus, the effect would be a lower population. This, then, is the theoretical mechanism behind the population projection: one should expect faster population growth rates leading to larger population and city size under closed city conditions (i.e. high population scenarios) whereas open city conditions should result in a slower, and therefore smaller, population growth.

It is worth noting that in the most recent United Nations city population predictions (United Nations, 2018), both Uganda and Rwanda only report one city with over 300 thousand residents (Kampala and Kigali), contrary to many of its larger neighbours such as Ethiopia, Kenya, or Tanzania, lending credence to the closed city assumption. However, the urban national policy of Rwanda, currently under development, envisions the reinforcement of secondary cities, which if successful could create the possibility of an open city within an urban hierarchy (to be sure, Kigali would continue, even then, as the main population and economic center of the country).

7.2.2 Land supply

As noted in chapters 4 and 5, how much land demand is realized (actually built) and where is a function of the spatial determinants of urban growth, which reflect the desired locations of urban agents, and of the urban system's capacity to supply the buildings. Estimates of this potential to supply built-up area are required by the urban growth model to allocate development. Under calibration conditions (as in chapter 5), this model of potential supply is known because the urban growth of the calibration period is known. When projecting urban development, the supply map may reflect the unplanned past (as in Upper Lubigi, Kampala, see chapter 4) or the planned trend (see for example the case of Rwampara, Kigali, Pérez et al., 2016).

A second element to consider, however, is the difference in total amount and per cell amount of urban growth in greenfield vs. redevelopment (intensification of existing urban areas). Table 7.1 summarizes total urban development for Kigali and Kampala, distinguishing between greenfield development and intensification of the existing urban fabric. One can see a tendency towards greater dispersed greenfield development (increasing quantity and percentage of greenfield development), which is likely the result of greater prosperity (greater stability from the perspective of security, possibly rising incomes). Furthermore, when estimating the average increase of built-up fraction for greenfield and intensification areas (excluding cells with 0.00 built-up fraction), one finds another striking difference: for the 2001-2016 period, greenfield development cells of Kampala increased on average 0.182 vs. an increase of 0.364 for the intensification areas; in Kigali, 2000-2015, these average increases were of 0.112 and 0.224 respectively. This is in line with theoretical expectations, since capital-to-land ratio is a decreasing function of accessibility, hence one should expect the more accessible existing urban areas to develop at greater pace than the periphery.

To generate models of potential supply of prospective scenarios, one first must consider the unplanned condition. The potential supply map was created by applying map algebra following equation 7.1:

$$SimDem_i = 2 \cdot AvgBUfr \cdot UnifRandom_i \tag{7.1}$$

Table 7.1 Urban growth per type in Kampala and Kigali (ha)

		Urban growth	
Period	Total	Intensification	Greenfield
Kampala			
2001-2010	6830	2781	4049
	100%	41%	59%
2010-2016	8267	2737	5530
	100%	33%	67%
Kigali			
2000-2009	1125	609	516
	100%	54%	46%
2009-2015	1471	708	763
	100%	48%	52%

Intensification: urban growth occurring in areas that already presented urban development in initial year. Greenfield: urban growth occurring in areas that had 0.00 built-up fraction in initial year.

with $SimDem_i$ the potential supply of cell i, $UnifRandom_i$ a random value in the range $[0.00, 1.00]$; the overall distribution of $UnifRandom$ is spatially random, and $AvgBUfr$ the average built-up increase per cell. The 2 in equation 7.1 must be introduced because the expected value of $UnifRandom$ is 0.50 so, in its absence, the average built-up increase would be halved relative to the reference data. This procedure is equivalent to what was implemented in chapter 4.

The discussion on greenfield vs. intensification urban growth and the need for an average per cell built-up fraction introduces a new requirement into the prospective simulation, relative to the models used in chapters 4 and 5: when considering unplanned scenarios, two potential supply maps must be used, one for greenfield development and another for intensification (because of the difference in the $AvgBUfr_i$ parameter); the increase of built-up fraction is a function of type (whether in the urban fabric or not) but not its location, which follows from the suitability score that combines the spatial determinants (derived in chapter 5). However, because location is determined in no small part by accessibility, if all land demand were allocated jointly, one should expect an unrealistic densification of central locations and hardly any greenfield development (contrary to the trends present in the data for previous periods). Therefore, for prospective simulation, land demand was also divided into two fractions, according to a percentage of expected greenfield development. One should note that, when densification is prevalent in a scenario, this greenfield fraction should be reduced and the average per cell increase of built-up fraction should be enlarged, relative to trend conditions of both cities.

Compared to unplanned conditions, the potential supply for planned

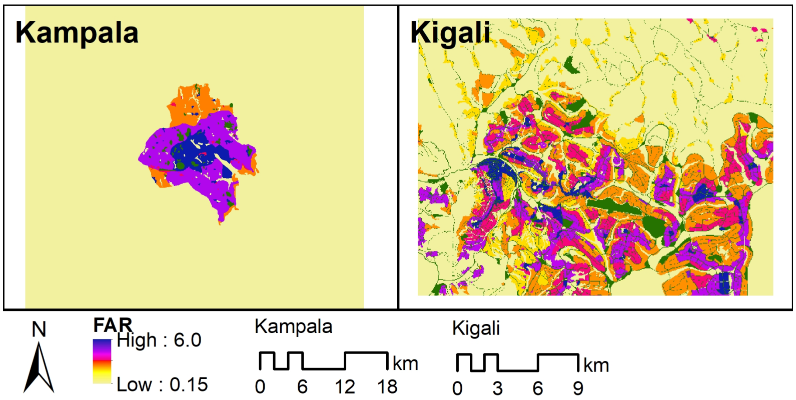


Figure 7.2 Floor-to-area ratio of Kigali Master Plan and interpretation of Kampala Physical Development Plan

scenarios is straightforward. For Kigali, the Master Plan (SURBANA International Consultants PTE Ltd., 2013) explicitly defines floor-to-area (FAR) ratios for residential and commercial land use zones – for industrial zones, this parameter was assumed equal to 1.00, since only maximum percentage of covered area is regulated; for protected areas, for which constructions are assumed minimal, the FAR was set to 0.15. Because the Kampala Physical Development Plan is strategic rather than normative in nature (ROM Transportation Engineering Ltd. et al., 2012), no such parameterization was available. However, land use zones between both plans were compared and the parameters of the closest equivalent from Kigali were used to generate an estimated FAR map for Kampala. These maps of FAR were adopted as potential supply; in effect, they call for intensification of central locations via multi-storey buildings, since FAR values of Kigali are large (generally greater than 1.00 for commercial and residential areas), as is shown in figure 7.2.

Because FAR values are relatively large and due to technological and social constraints, especially in terms of income of urban agents, the land use plan of Kigali has been considered unrealistic (Watson, 2014) and by extension, similar criticism is possible of the potential supply model for Kampala. Indeed, previous experiments (Pérez-Molina et al., 2016) already supported the conclusion that the available development potential far exceeded the needs of Kigali, i.e. that the FAR had been set to excessively large values. Following this previous work and to increase the realism of the planned scenarios, an assumption was made that potential supply of the land use plans would only be partially realized – between 50% and 100% of the regulated FAR would actually be built, following a spatially random pattern.

7.2.3 Rainfall and flooding

The relations between rainfall and flooding can be extraordinarily complex, as flooding depends not only on the magnitude of rainfall events but also on antecedent precipitation, characteristics of the landscape, and spatial variation of all of these variables (Turkington, 2016). Compounding the problem, climate change will likely introduce variations in expected triggers of natural hazards, of unknown effect. Even in cases with good data available, exploring these multiple dimensions is a challenging task (Turkington, 2016). In Sub-Saharan Africa, to the inherent problems of the phenomenon one must add the lack of comprehensive measurements of both rainfall and river flows, which preclude much of the traditional calibration exercises of flood modeling.

How, then, can one deal with flooding and its effect on prospective urban growth? Much of the evidence and methods applied for scenario construction were already described in chapter 3: (1) Areas exposed to flooding impose additional costs to urban agents and, because of this, they will be undeveloped relative to areas with similar accessibility. However, as population increases, urban agents will be increasingly willing to accept exposure to flood when compensated by greater accessibility (Frame, 1998). (2) Peak discharge, as a proxy of flooding, does have a measurable deterrent effect on urban growth; but this effect is very small when compared to the growth triggered by population increase and should fade as population continues to grow. A corollary of this conclusion, the main finding of chapter 3, is that little to no impact of flooding on urban patterns should be expected for smaller populations because most of the exposed area will not have been yet developed. In consequence, the feedback effect between flooding and urban growth has only been explored for Kampala, the larger metropolitan region, as widespread effects are still not expected for Kigali. (3) While information is scant, there is sufficient evidence to derive maximum yearly rainfall events for selected return periods (this chapter adopts the estimates of Chogyal, 2012) and to synthesize events using intensity-duration-frequency curves (Fiddes and Forsgate, 1974) for Kampala. (4) The constraint on urban development caused by flooding was detected by assuming the maximum rainfall of the past five years causes the constraint; in effect, this is equal to assuming urban agents remember the maximum flood that has occurred over that period.

To simulate the 2016-2030 period incorporating a feedback effect between flooding and urban growth, a first assumption is made relative to the results of chapter 5: recall a wetlands factor was calibrated and interpreted as reducing suitability for urban development in locations occupied by wetlands; the reasons for this included increased exposure to flooding, soils subject to subsidence, protections – even if inefficient – of the land use regulation system, among others. The first assumption is, then, to accept exposure to flooding as the main cause of the constraint posed by wetlands. In consequence, the map of wetland areas can be substituted by the map of flooded areas, in turn affecting urban growth

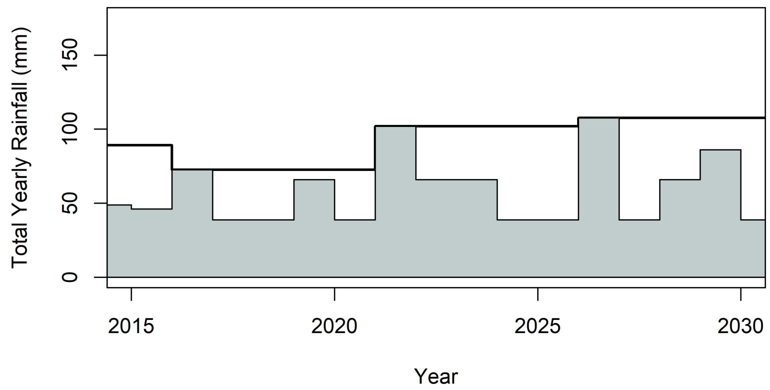


Figure 7.3 Simulation of rainfall events for prospective scenarios of urban growth

allocation each year.

The need for a flood map every year simulated further complicates matters, as flooding is estimated using the event-based *OpenLISEM* model. Events, as noted, can be simulated for any given daily rainfall and the daily rainfall of a given return period can also be determined. Yet what event should be assigned to each year of the prospective simulation? The following assumptions were made for the 14 year simulation period: events of return period equal to 1 were exceeded every year of the simulation period, events with return period equal to 2 were exceeded half of the time (seven years), events with return period equal to 5 were exceeded three times, events with return periods equal to 10 and 20 were exceeded one time only. If one were to select one event with return periods of 5, 10, and 20 years, four events with return periods of 2 years, and seven events with return periods of 1 year, all of the assumed constraints would be met save the 10 year return period event being exceeded twice. This set of events was randomly arranged; the result of this sorting is shown as a gray area in figure 7.3. Because the corresponding event to every year in the simulation period, as well as for preceding years (see chapter 3), is known, it is a simple matter to determine the maximum event of the past five years (shown as a black line in figure 7.3).

The total yearly rainfall events reported in figure 7.3 were used to synthesize hyetographs with 15 min intervals using the intensity-duration-frequency analysis already described in chapter 3. A simplifying assumption, that this five year maximum event occurred in each simulated period, was adopted (strictly speaking, one should calculate a flood map for each year using that year's corresponding rainfall and choose the maximum flood of the five year period as the one triggering behavioral changes of urban agents, but this is a cumbersome process that adds little in the way of meaningful variance, given the limited sensitivity of flood estimates to patterns with relatively large built-up fractions vis-à-vis the rainfall events).

7.3 Urban growth: planned and unplanned scenarios of Kampala and Kigali

In this section, the possibilities identified as critical uncertainties were combined into scenarios, with the aim of generating alternative development paths uniquely associated to a specific variations that contribute to explore the focus question *how could urban growth and flooding be influenced through land use planning?* Two series (A for Kampala, B for Kigali) of seven scenarios (1-3 for unplanned potential supply, 4-7 for planned potential supply of urban land) were constructed and compared to each other, eliciting knowledge on a set of future possibilities of Kampala and Kigali, and how land use planning, flooding, and urban growth may factor into this set. Table 7.2 summarizes the combination of critical uncertainties that defines each scenario.

Meta-narratives for scenarios 1-3 and 4-7 are presented as a means of discussing the salient characteristics of each scenario; a discussion then compares both groups in the context of the focus question.

7.3.1 Unplanned prospective scenarios narratives

Circa 2015, the institutional environment underpinning the land use planning system, and more broadly public institutionality, looses legitimacy. Regulatory constraints break down; urban growth patterns become controlled by land market dynamics reflecting urban agents' preferences and the urban system's capacity to materialize them.

As part of this generalized collapse, the city becomes a relatively safe haven in terms of livelihood, driving up immigration but without any material prosperity arising. Population growth follows the high population growth assumption and the urban system is capable of accommodating it without changes to urban agents' preferences as expressed in the circa 2010-circa 2015 period (scenario 1); alternatively, in the short run, the urban system proves incapable of supplying the required urban land: the city densifies under closed city assumptions (high population growth remains, scenario 2; densification is achieved by reducing the percentage of land demand allocated as greenfield and by doubling the average per cell increase of built-up fraction) or the city looses excess population (density preferences remain, low population growth, scenario 3).

7.3.2 Planned prospective scenarios narratives

Circa 2015, the institutional environment crosses a critical threshold that provides the social legitimacy and political incentives to strongly enforce land use regulation. Land use planning regulations envisioning

Table 7.2 Scenario definitions

		Kampala		Kigali	
		High pop. growth	Low pop. growth	High pop. growth	Low pop. growth
Unplanned	Baseline densification	1A	3A, 2C*	1B	3B
	Increased densification	2A, 1C*		2B	
Planned	Partial implementation	4A, 3C*	5A	4B	5B
	Full implementation	6A	7A	6B	7B

* indicates scenarios for which a feedback between urban growth and flooding was implemented. Baseline densification: based on 2010-2016 for Kampala and 2009-2015 for Kigali; increased densification: built-up fraction increase was doubled relative to the baseline period and % of land demand allocated as greenfield was reduced. Full implementation (of land use plan): the envisioned FAR is totally implemented; partial implementation: only between 50% and 100% (following a spatially random pattern) of the total FAR potential is actually realized (in practice, this equates to a reduction of development density).

a multi-storey building-based densification of central locations coupled with a strong constraint on the amount of permitted development in peripheral areas are available for implementation.

Urban agents' preferences are thus constrained by regulation. The urban system is fully capable of building the envisioned structures and urban agents have the financial capability to occupy them (scenarios 4 and 5) or the system and urban agents can only partially take advantage of the opportunities presented by the regulation while being forced into following its constraints (scenarios 6 and 7).

The city continues to attract immigrants at a rapid pace (high growth population, scenarios 4 and 6) or at a slower pace (low growth population, scenarios 5 and 7).

7.3.3 Prospective scenario results: a discussion

The simulated prospective scenario results for the target year 2030 are shown in figure 7.4 for Kampala and figure 7.5 for Kigali. The bottom

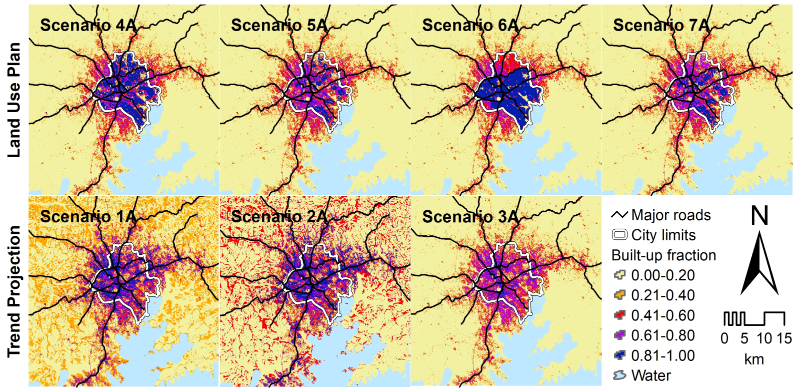


Figure 7.4 Prospective urban growth scenarios of Kampala, 2030. Scenarios 1A, 2A, 4A, and 6A correspond to high population growth, scenarios 3A, 5A, and 7A correspond to low population growth. Scenarios 2A, 6A, and 7A correspond to increased density, relative to baseline conditions.

row of each figure (scenarios 1-3) correspond to unplanned land supply and the top (scenarios 4-7), to planned land supply.

A first question regarding all scenarios is, can the environment accommodate the land demand associated to high population growth? The answer is unequivocally yes, land demand was fully allocated in all scenarios but two: scenario 6A and scenario 2B. These two become important as bounding conditions:

Scenario 6A represents a partial implementation of the **Kampala** land use plan under the high population growth condition. Because the simulation cannot totally accommodate demand, the scenario result strongly suggests the urban system will need to take full advantage of the densification opportunities opened up by land use regulation; this in turn is problematic because there are at least two constraints in the

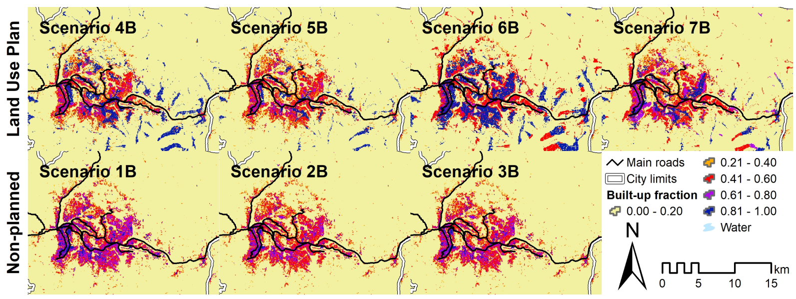


Figure 7.5 Prospective urban growth scenarios of Kigali, 2030. Scenarios 1B, 2B, 4B, and 6B correspond to high population growth, scenarios 3B, 5B, and 7B correspond to low population growth. Scenarios 2B, 6B, and 7B correspond to increased density, relative to baseline conditions.

short-run to such a development: there are technical limitations (building multi-storey buildings requires the use of technology in lieu of manpower in a context of very low labour costs, which means this technology is at present very likely underdeveloped) and societal limitations, specifically Kampala's population is generally poor; most residents cannot afford apartments in multi-storey buildings. Furthermore, such generalized poverty causes systemic problems to the real estate market, in part manifested through widespread informality. In sum, it is very unlikely scenario 4A will come to pass and even scenario 6A may be optimistic in assuming a 50% or more development according to the plan.

Scenario 2B represents an unplanned development of **Kigali** if the city were to densify. One should note that, in the face of increasing population, densification is a theoretically expected result of a closed city growth (Brueckner, 1987); thus, scenario 2B is more likely than scenario 1B. Why would scenario 1B be more feasible, computationally, than scenario 2B? First, one must stress that Kigali cannot continue to develop with only 52% of total demand as greenfield development unless the land use plan is implemented at least partially (scenarios 4B-7B) – for scenarios 1B-3B, the 67% of land demand, the percentage of Kampala 2010-2016, was allocated as greenfield development. By doubling the per cell average increase of built-up fraction (as in scenario 2B, relative to scenario 1B), the model reduces the amount of greenfield development at each iteration. This eventually mounts up to the point that there is not sufficient built-up area for land demand under intensification to be fully allocated. Ultimately, then, comparing scenarios 1B and 2B one may conclude the historically compact building patterns of Kigali can only continue into the future if single storey buildings are substituted by multi-storey buildings, as envisioned by the land use plan, and subject to the same constraints discussed for Kampala in the preceding point: thus, Kigali is very likely to sprawl more in the near future.

What can be learned from an examination of the resulting urban patterns for 2030?

For **Kampala**, as one should expect, (1) scenarios of unplanned trend conditions (scenarios 1A-3A) sprawl (cause greenfield development in the periphery) much more than planned scenarios (scenarios 4A-7A) and (2) high population growth scenarios (1A, 2A, 4A, and 6A) result in more development than low population growth scenarios (3A, 5A, 7A); this takes the evident (in figure 7.4) form of much greater greenfield development in the unplanned scenarios and much more densification of the urban core for the planned scenarios. In this sense, the most balanced patterns, and likely the best from a broad sustainability viewpoint, correspond to low population growth conditions; thus, paradoxically, the likely best approach for Kampala passes through a national policy to reinforce secondary cities in Uganda.

Unexpectedly, for **Kigali**, the unplanned scenarios (1B-3B) produce patterns that seem to be clearly more compact than the planned scenarios (4B-7B). Unplanned conditions are a reflection of trend characteristics, which in Kigali already incorporate the constraints posed by the land use plan to sprawl; since the city is still relatively small, land demand has not yet (by 2015) pressured the land market towards widespread densification of central locations (which may or may not be feasible). Since scenarios 1B-3B embody these constraints, they are compact as should be expected. What is more surprising is that the land use plan promotes dispersion of urban growth, particularly the emergence of new urban cores in the east and south of the Kigali metropolitan region. Ultimately, the explanation lies in the land use zoning scheme (figure 7.2) which plans for a much larger Kigali than what the projections suggest will be materialized. The advantages the regulation gives to certain peripheral locations, however, may result in these relatively inaccessible areas being occupied despite more central locations being still available.

How do the land use planning instruments affect flooding? In chapter 5, the wetlands factor was identified as a meaningful determinant of urban growth. It was also much more important for Kigali than for Kampala, in part reflecting regulatory constraints imposed in Kigali at least since circa 2008. This area of wetlands corresponds to either ecosystems of permanently flooded vegetation or to floodplains and, in both cases, to the area most exposed to recurrent flooding. Therefore, to indirectly evaluate the impact of regulation on potential flood risk, the amount of built-up area within the wetlands was quantified in table 7.3.

Table 7.3 Built-up areas (ha) potentially exposed to flooding in Kampala and Kigali for projected scenarios, 2030

	Kampala		Kigali	
Scenario	Area (ha)	%	Area (ha)	%
Unplanned scenarios				
Scenario 1	3520	8.3%	832	18.5%
Scenario 2	3288	7.7%	493	11.0%
Scenario 3	4367	10.2%	596	13.3%
Planned scenarios				
Scenario 4	3100	7.3%	811	18.1%
Scenario 5	3100	7.3%	811	18.1%
Scenario 6	5243	12.3%	530	11.8%
Scenario 7	3100	7.3%	518	11.5%
Wetlands total	42661	100.0%	4488	100.0%

The built-up areas potentially exposed to flood do not show any systematic relation to land use plan enforcement. For Kampala, scenarios 3A and 6A correspond to the largest proportion of wetland areas developed (3A an unplanned scenario, 6A a planned scenario, all others approximately the same at around 7.5% of wetland area developed). The explanation for scenario 6A is likely related to the fact that it could not accommodate the entire land demand and, thus, expanded to the greatest extent possible. The feasibility of developing at high density (i.e. multi-storey buildings) may be, thus, linked to flood mitigation. Scenario 3A, in turn, can be explained because scenarios 1A and 2A sprawl in part following the road system, spreading development away from wetland valleys in Kampala's hinterland; paradoxically, by being more compact, scenario 3A presents a pattern that develops a greater proportion of the environmentally sensitive wetland areas near and within the city of Kampala proper. What is less clear is why scenario 3A would result in more potentially exposed areas than scenarios 4A, 5A, and 7A. A possibility is that land use plans protect some wetland areas in central locations (lower values of FAR are assigned to them and others are directly blocked), although one would not expect too large an effect from this characteristic, since the constrained area is limited.

For Kigali, scenarios 1B, 4B, and 5B result in higher percentage of wetland areas developed: scenarios 4B and 5B correspond to the full implementation of the plan; scenario 1B is the unplanned development under trend conditions. Since scenario 1A should sprawl more than scenarios 2B and 3B, it is consistent with the scenario design that scenario 1A should cause more development in the wetland areas. Explaining why full implementation of the plan results in more wetland areas being developed is less straightforward: the wetland area being occupied is probably around central locations with large FAR values – and in the partial implementation scenarios (6B and 7B) these areas have smaller neighborhood factor values (as they accommodate less development compared to the full implementation).

Summing up, Watson (2014) has argued correctly that the land use plan of Kigali (and the interpretation here adopted for Kampala, based on it) is not appropriate for the social and environmental context of the city. Even on its own terms (of promoting compact development), the Kampala Physical Development Plan is positive but the Kigali Master Plan may have counterproductive effects, by creating an actually sprawling pattern. The implementation of the land use plans in these two cities is very unlikely to reduce flood risk, since they promote very high built-up fractions in central locations (a reduction of infiltration and, in consequence, the generation of more runoff) and these do not necessarily preclude sprawl (resulting in more exposed development at other locations). The key to mitigate flooding is to stringently protect wetlands from development; runoff reduction is unlikely to be achieved without infrastructure interventions. Current land use plans do not incorporate these environmentally proactive strategies with enough strength.

7.4 The feedback between flooding and urban development in Kampala

This section reports scenarios designed to explore the focus question *how do urban growth and flood processes interact?* Three scenarios (1C-3C), all for the metropolitan region of Kampala, were implemented to operationalize a dynamic feedback effect between urban growth and flooding. As noted in section 7.2, in the urban growth model, the flood map resulting from a rainfall event substitutes the wetlands factor. Three events, the hyetographs of which are reported in the appendix, were used – because, as in chapter 3, the largest event of the past five years is thought to influence the locational behavior of urban agents.

Computationally, each scenario executes in succession the following programs, for every year between 2017 and 2030: (1) the cellular automata urban growth model as in section 7.3, substituting the flood map for the wetlands map, to produce the built-up fraction projected map, (2) instructions to allocate bare soil and vegetation fractions, as in chapter 4 (in brief, the land fraction not built-up is divided into bare soil and vegetation, proportionally to the fraction of these land covers in the baseline year), (3) instructions to mask the land cover fraction maps (built-up, vegetation, bare soil) to the extent of the flood model, as reported in chapter 3, (4) a PCRaster script to update the flood model inputs determined by the land cover maps – the built-up fraction and vegetation fraction maps but also Manning's n coefficient map, the leaf area index map, etc., (5) the *OpenLISEM* flood model with the updated inputs and the rainfall event corresponding to each year, as shown in figure 7.3, (6) instructions to reclassify the flood depth map, taking any cell with depth over 15 cm as flooded and all others as non-flooded and to change the extent of the flood map from the catchments of the flood model to the full extent of the urban growth model (see chapter 5). Since PCRaster is not very flexible when dealing with changes of extent, these instructions (from the urban growth model extent to the catchment extent and vice versa) were implemented using ArcGIS 10.5. All other spatial analysis was run with the PCRaster Python Extension, except for the *OpenLISEM* flood model and the programs used to convert .map files to .asc files (for use in ArcGIS) and .asc files to .map files (results from ArcGIS processes), all of which were run as programs called within the Python code.

Using this approach, three scenarios were developed. Scenario 1C adopts unplanned conditions and allocates the high growth population land demand under densification conditions (its parameters are equal to those of scenario 2A). Scenario 2C also adopts unplanned conditions but it allocates land demand under low population growth (with parameters equal to those of scenario 3A). Finally, scenario 3C adopts the planned conditions of potential supply with high growth population and partial implementation; its parameters are equal to scenario 6A.

The results of the simulation are shown in figure 7.6. Already examining the overall patterns, one may see the first differences introduced by substituting the wetland maps by the flood map: for unplanned scenarios, the amount of desirable (suitable) area is greater when implementing the feedback because the extent of flooding is much smaller than the total wetland area. The consequence is the emergence of a much more dispersed pattern for both scenarios 1C (relative to 2A, its non-feedback equivalent) and 2C (relative to 3A, its non-feedback equivalent). In both cases, new and relatively dense centralities emerge in the periphery and remote areas (i.e. far from Kampala's center) are developed to a large extent, such as Entebbe (south). Comparing scenarios 1C and 2C, as expected, scenario 1C shows much more development in the periphery – because it allocates a much greater land demand. It is also notable that wetland areas in both scenarios are nearly fully developed in central locations.

In contrast to scenarios 1C and 2C, scenario 3C is much more compact and wetlands are much better preserved (the Lower Lubigi and the Nalukolongo wetlands, centrally located, are clearly visible). Development follows a more ordered pattern along main roads and at likely greater densities than the unplanned scenarios. Looking at the overall pattern, the land use regulation seems successful in containing urban sprawl. When considering the Upper Lubigi subcatchment, as an example of division-level (sub-city) scale, some benefits are also shown: while practically the entire sub-catchment is developed, the planned scenario (3C) results in lower intensity of development around the main and secondary drainage channels, relative to scenarios 1C and 2C.

The main flood model outputs are shown in figure 7.7 and table 7.4. The results show that, overall, there are many flood outcomes that are not affected by the critical uncertainties which determine the scenarios: the flood patterns derived for all three scenarios are nearly identical (see figure 7.7); they also highlight one of the problems introduced by substituting wetlands with flood, that the area actually flooded even by relatively large rainfall events (with a recurrence period of 20 years) is a small portion of the metropolitan region. In table 7.4, one may see that in addition to total flooded area, infiltration is also identical for all three scenarios. However, and crucially, the amount of built-up area flooded is substantially smaller for scenario 3C (high population growth, planned) than for scenario 1C (high population growth, unplanned); the exposed built up area of scenario 3C is even smaller than for scenario 2C, despite allocated land demand being 30% smaller.

The judgement of land use regulation as essentially ineffective, which follows from the analysis of section 7.3, may have been unwarrantedly pessimistic. However, it is important to stress that the potential effects of regulation should be more visible in larger systems (as the constraints to development of areas exposed to flooding, inherent to urban agents' locational preferences, should fade as population increases; this sharpens the difference between regulated and unregulated urban patterns). Kampala in 2016 was already a larger city than what Kigali is expected to become by 2030, even under high population growth. Furthermore, as noted in section 7.3, the effects of the land use plans in terms of promoting compaction or dispersion of urban growth are different for Kampala (essentially, regulation incentivizes compact growth) than for Kigali (for which the land use plan seems to produce development and even new core areas in peripheral locations). Both of these conditions affect the potential of regulation to limit exposure to flooding.

Synthesizing, the implementation of a feedback effect between urban growth and flooding visibilizes the positive effects of regulation on flood risk mitigation. Consistent with the results previously detected in chapter 3, the impacts of flooding on land patterns detected through prospective

Table 7.4 Flood impacts on built-up land cover for prospective scenarios of Kampala with feedback between flooding and urban growth, 2030

	Scen. 1C Trend Intensif.	Scen. 2C Trend Low Growth	Scen. 3C Plan
Flooded area over 15cm (ha)	1598.5	1599.3	1599.0
Infiltration (mm)	81.4	81.4	81.4
Built-up area total (ha)	16682.9	14874.9	15830.5
Built-up flooded over 15cm (ha)	553.4	423.1	407.7

scenarios are noticeable but modest. The impacts on runoff generation of land use regulation are essentially non-existent because the population growth simulated results in large built-up fractions regardless of the patterns of urban growth (with the caveat, noted in subsection 7.2.3, that the relations between different physical processes which ultimately explain floods are complex and dependent on initial conditions often unknown). Flood patterns are also nearly identical regardless of potential supply (whether planned or unplanned) and population (high or low population growth). However, the urban agents' (modest) desire to avoid areas exposed to flooding reinforces the constraints proposed by the regulation, which cumulatively result in less built-up area flooded.

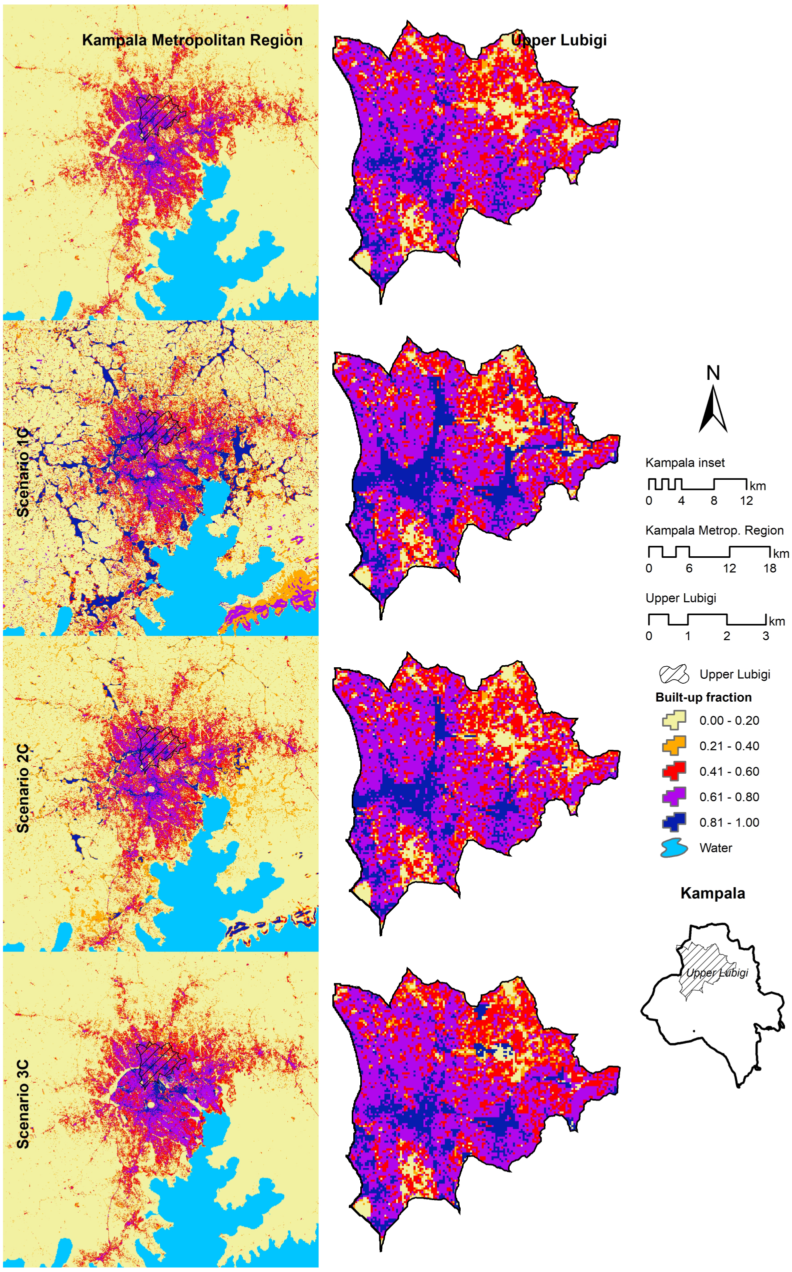


Figure 7.6 Prospective urban growth scenarios of Kampala incorporating feedback effects of flooding and urban growth, 2030. Scenarios 1C and 3C correspond to high population growth, scenario 2C corresponds to low population growth. Scenario 1C corresponds to increased density, relative to baseline conditions.

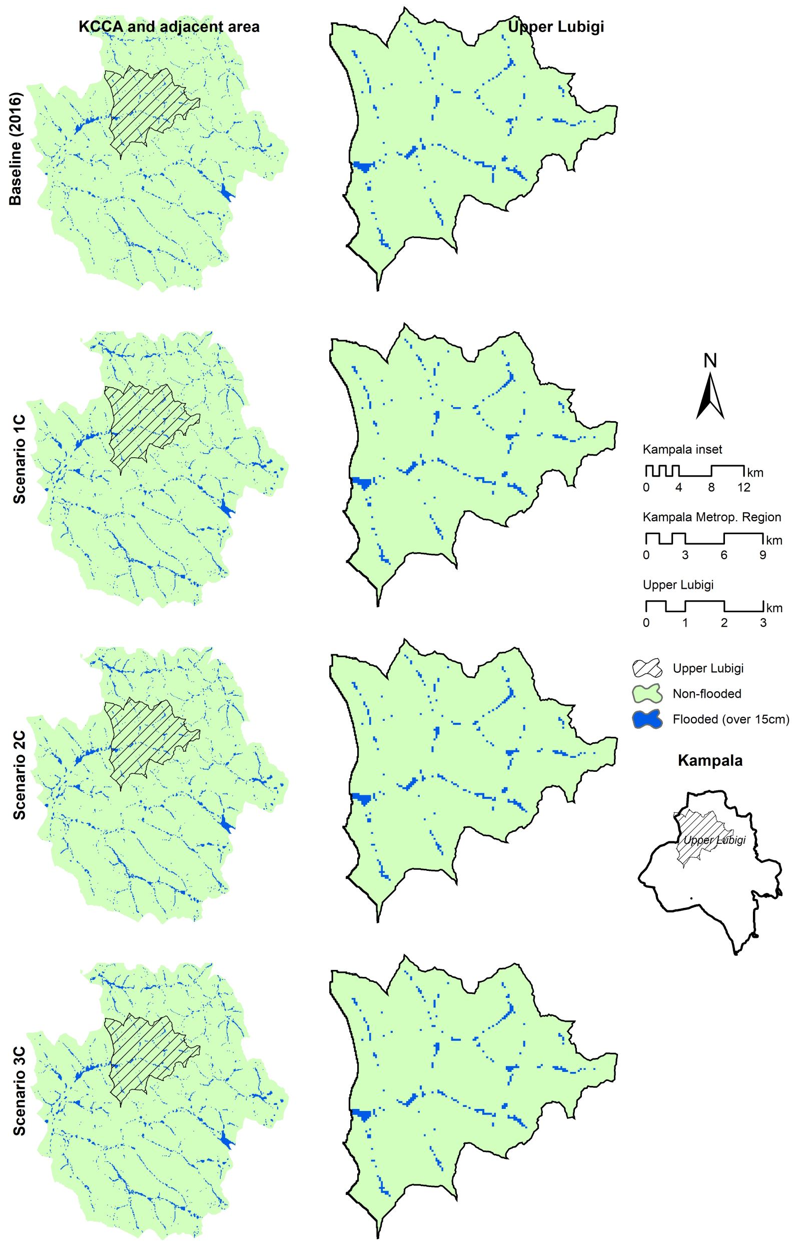


Figure 7.7 Prospective flooding scenarios of Kampala incorporating feedback effects of flooding and urban growth, 2030.

7.5 Appendix

Table 7A.1 Zoning and interpretation of the Kampala Physical Development Plan

Zone	Interpretation	Parameterization
City Center	Equivalent to C4 (Regional Level Commercial District)	FAR = 6, Cov = 0.7
Business, Community Service, Commerce (secondary center)	Equivalent to C3 (City Level Commercial District)	FAR = 3, Cov = 0.7
Tourism and recreation	Equivalent to C3A (Historic, Cultural, Tourism and Recreational)	FAR = 2.4, Cov = 0.8
Central Residential Zone	Equivalent to Medium Rise Residential District (R3)	FAR = 1.8, Cov = 0.5
Inner Area Residential Zone	Equivalent to Low Rise Residential District (R2)	FAR = 1.4, Cov = 0.6
Peripheral Residential Zone	Equivalent to Single Family Residential District (R1)	
Lakefront	Equivalent to Single Family Residential District (R1)	FAR = 0.8, Cov = 0.4
Urban park, natural forest reserve, sports facilities	Protected (assumption)	FAR = 1.0, Cov 0.15
Public facility	Protected	FAR and Cov = 0
Existing industrial areas	Equivalent to General Industrial District (I2) (FAR assumed)	FAR = 1, Cov = 0.6

Table 7A.2 Scenario parameterization

			Population				Average growth per pixel	
Scenario	Year	Path	Density	Projection	Land demand	Greenfield %	Greenfield	Intensif.
Kampala								
Base	2016		128.7	2706790		67%	0.182	0.398
Scen. 1A	2030	High growth		5505707	42778	67%	0.182	0.398
Scen. 2A	2030	High growth		5505707	42778	60%	0.364	0.796
Scen. 3A	2030	Low growth		3870785	30075	67%	0.182	0.398
Scen. 4A	2030	High growth		5505707	42778	60%	FAR (full)	
Scen. 5A	2030	Low growth		3870785	30075	60%	FAR (full)	
Scen. 6A	2030	High growth		5505707	42778	60%	FAR (partial)	
Scen. 7A	2030	Low growth		3870785	30075	60%	FAR (partial)	
Kigali								
Base	2015		251.6	951359		52%	0.112	0.352
Scen. 1B	2030	High growth		2213868	5018	67%	0.112	0.352
Scen. 2B	2030	High growth		2213868	5018	67%	0.224	0.704
Scen. 3B	2030	Low growth		1568247	2452	67%	0.112	0.352
Scen. 4B	2030	High growth		2213868	5018	52%	FAR (full)	
Scen. 5B	2030	Low growth		1568247	2452	52%	FAR (full)	
Scen. 6B	2030	High growth		2213868	5018	52%	FAR (partial)	
Scen. 7B	2030	Low growth		1568247	2452	52%	FAR (partial)	

Density: residents per ha, land demand: ha

Table 7A.3 Rainfall hyetographs of simulated events

		Intensity (mm/h)		
		T = 20	T = 10	Year 2016
0	30	7.72	7.32	2.81
45	60	5.59	5.31	4.07
60	75	7.75	7.35	5.64
75	90	11.76	11.16	8.56
90	105	20.87	19.80	15.20
105	120	51.59	48.95	37.58
120	135	155.87	147.90	113.53
135	150	51.59	48.95	37.58
150	165	20.87	19.80	15.20
165	180	11.76	11.16	8.56
180	195	7.75	7.35	5.64
195	210	5.59	5.31	4.07
210	240	7.72	7.32	5.64
Daily total		107.5	102.0	78.3
Event total		91.6	86.9	66.7

8 Recapitulation: urban growth, flooding, and spatial planning

The starting point of this dissertation was a problem – the cities of Kampala and Kigali suffer, to different degrees, recurrent flooding that negatively affects the quality of life of their inhabitants – in a context (Sub-Saharan Africa) and belief in a solution: effective land use planning can contribute to mitigate flood risk and, thus, improve people's lives. Developing this proposition necessarily led to qualifications of all three elements: by acknowledging that the context is diverse (there are important differences between Kampala and Kigali in particular and among the cities of Sub-Saharan Africa in general), that the problem takes different forms in space and time – e.g., Kigali's present flood problems are localized whereas Kampala's are more widespread, a consequence of institutional differences but also of varied characteristics such as: overall population, slope and soil infiltration patterns, ecosystems, etc. –, and that these qualifications, all taken together, contribute to create multiple plausible paths for the future of these cities. These paths diverge in terms of how effective land use planning can be in shaping flood risk and, for paths along which land use planning is effective, in terms of whether the role of such regulation is positive or negative. The results of this dissertation, then, should be thought of as a description of the conditions under which disparate urban patterns develop.

How do urban growth and flood processes interact? Theoretical models of urban location (Frame, 1998) propose that, all else held equal, areas exposed to flooding will present lower capital-to-land ratios (less buildings) than unexposed areas. Within this tradition, locations exposed to flood are determined by critical risk levels associated to a hazard with a defined probability of occurrence (i.e. to the magnitude of the hazard, since less probable events are larger). This critical level is a negative function of commuting costs: urban agents will be more willing to accept exposure to flooding if compensated by greater accessibility. Crucially, as the overall population of the city increases, the critical level is reduced (areas previously undeveloped because they were exposed to flooding are occupied by urban agents).

The conclusions of this theoretical model were tested for Kampala (2001-2016) by means of a structural equations model (these results are reported in chapter 3): a causal chain was postulated, triggered by population growth which resulted in greater urban growth and larger urban areas; larger urban areas, in turn, produced greater flooding that reduced urban growth. The marginal effect of the flooding proxy variable on urban growth was found to be statistically significant but small, especially in relation to the direct effect of population on urban growth (which is three to ten times larger). One must conclude the constraining effect of recurrent flooding is not large enough to mitigate flood risk by itself, but it could (at certain stages of a city's development) open a window of opportunity for policy actions by delaying urban development in areas exposed to flood.

This conclusion was bolstered by the cellular automata urban growth model's calibration outcome, reported in chapter 5. For both Kampala and Kigali, wetland areas were found to be inversely associated to urban growth. The factor was 3.5 times larger in Kigali than in Kampala (in the smaller city than in the larger, consistent with the idea that the constraint posed by flooding fades as a city becomes larger); indeed, this may contribute to explain why flooding in Kigali is less widespread than in Kampala. The result is important because, unlike the structural equations models reported in chapter 3 (which analyze data aggregated to the sub-catchment level), the cellular automata model describes variation between locations. A number of reasons contribute to explain why wetland areas are less likely to have been developed in Kampala and Kigali between 2000 and 2015: both cities (Kigali more efficiently) made efforts to protect natural wetlands from development and these areas are also geotechnically unsuitable for urban development. Yet recurrent flooding occurs mostly within or bordering these natural wetland areas (and where it doesn't, it happens in areas that would be wetlands had they not been drained for development). It is very likely that exposure to flooding is one of the reasons why urban agents avoid the wetland areas of Kampala and Kigali.

These findings were used to design prospective scenarios of urban growth for Kampala, 2016-2030 (reported in chapter 7). One set of scenarios made use of an integrated version of the model calibrated in chapter 5 (this integrated model had already been outlined in chapter 4) to implement a feedback between urban growth and flood patterns. Briefly, the wetlands constraint was interpreted as the consequence of recurrent flooding and for this reason the wetlands factor of the urban growth model was substituted with the flood map. This flood map, in turn, was updated for every time period simulated to reflect changes from the larger urban area and the simulated rainfall. The ensuing built-up land cover patterns clearly showed the most accessible wetland areas, including the recurrently flooded area adjacent to the main drainage channels, were fully developed. This expected result can be explained because flooding is already, at present, less important than accessibility in determining the urban patterns of Kampala.

Regarding Kigali, similar dynamics are unlikely to play out in the short run (in the long run, the non-linear nature of climate change and possible drastic population displacements make such speculation pointless unless purposefully tested through simulation). Kigali already is using land use regulation to constrain urban development in protected areas, among them wetlands (see chapter 6). In addition, Kigali is still expected to present a smaller population by 2030 than Kampala in 2015, even under rapid population growth. There should be sufficient available urban land for Kigali to develop without large increases of current flood risk – barring institutional collapse or other long run changes –, a judgement supported by the prospective scenarios developed for Kigali in chapter 7

To what extent can land use planning contribute to mitigate urban flood risk? Land use planning systems are regarded as generally ineffective and inefficient in Sub-Saharan Africa, a legacy of colonial rule even though decades have passed since independence (African Planning Association, 2014). Bucking the regional trend, Kigali has recently become known as an exception of order and strict implementation of the urban regulation (Goodfellow, 2013a). Can and should the example of Kigali be followed by other cities of Sub-Saharan Africa? What effects could one anticipate from the extension of such systems to other cities?

The causal effect of regulation on urban growth patterns for Kigali (2000-2014) was determined in chapter 6. A difference-in-differences model was applied to data structured as a two period panel of repeated measurements for a large set of locations; the variables measured were land cover (a binary variable of built-up/non-built) and the neighborhood effect identified in chapter 5; the model also controlled for spatial autocorrelation and panel fixed effects. It follows the general structure of spatial statistical difference-in-differences models applied to land cover patterns (Dempsey and Plantinga, 2013; Kline et al., 2014) extended to incorporate the interpretation of spatial effects within the causal framework, as proposed by Kolak (2017). The derived difference-in-differences estimators quantified a constraint, caused by the land use regulation, of between 0.023 and 0.030 – an apparently small effect until one considers the average built-up fraction of the protected area was 0.047 in 2014. In sum, the stringent application of land use regulations over the past decade (at least since circa 2008, according to Goodfellow, 2013a) did result in a measurable and important reduction of urban development.

When compared to other cities in Sub-Saharan Africa, Kigali stands out as an exception, in particular when compared to Kampala. Both cases, as was argued in subsection 1.4.2 from the work of Goodfellow (2013a), share a pysical and social context, including in particular very closely linked political elites. Why, then, does the land use system of Kigali function so much better than that of Kampala? Goodfellows's (2013a) answer relies on the idea of political bargaining environments of the elites exercising power: in Kampala, national politicians derive legitimacy from a populist approach, "often employing a rhetoric of intervening to 'protect' the poor from corrupt local government actors" (Goodfellow, 2013b, p. 46); Ugandan politicians often find value in fostering hostility, among certain socio-economic groups, against other parts of the state (the success of such tactics also is related to the greater ethnic diversity of Uganda, relative to Rwanda; Goodfellow, 2013b). This leads to a situation in which circumventing the regulation is possible because disorganization provides opportunities for Uganda's political elites. In contrast, the legitimacy of the national government in Kigali derives from its credibility as a law-enforcer, originally from stabilizing the country in the aftermath of the 1994 genocide; furthermore, and contrary to Uganda, the current political elites of Rwanda formed in exile: they have less developed relations with local actors, and thus depend less on them to maintain power. However, a very relevant exception has been detected in the literature: the coincidence of interests between developers of multi-storey buildings and political elites (Goodfellow, 2018) as well as politicians, e.g. lawmakers, often being property owners, which has contributed to weak land taxation (Goodfellow, 2017); often such conflicts of interest allow Rwandan political elites to use formal rules to favor powerful business interests, as well as themselves, at the expense of the poor. Even after accounting for these contradictions, support for political elites in Rwanda hinges on credibility for efficiency in achieving goals, first of security and by extension of economic development; because credibility is at stake, formal adherence to rules is of great importance, particularly urban rules because Kigali has become a showcase for national development (Goodfellow, 2018). Political elites in Rwanda ensure, through strong top-down accountability, that the rules are followed as a strategy to maintain the legitimacy which supports their exercise of power (Goodfellow, 2013b).

In addition to Goodfellow's ideas, it is also important to note that Kigali is a much smaller city than Kampala (one can view Kigali as an earlier stage of a Sub-Saharan Africa 'meta-city' and Kampala as a later stage; see subsection 7.2.1 for a use of this idea in understanding uncertainty associated to population projections); when considering environmental problems in similar physical contexts, city size matters: in particular, as was discussed for the relation between urban growth and flooding, the constraints posed by recurrent flooding tend to fade as cities grow in terms of population. What is true of flooding constraints is also true for regulatory constraints (in practice, both imply additional costs to urban agents), meaning pressure to evade the regulation is less for smaller cities.

Efficiency in the implementation of rules is necessary, yet not sufficient, for land use planning to contribute to flood risk mitigation. Land use regulation must also be effective, in the sense of nudging urban patterns into becoming broadly more sustainable. Watson (2014) has rightly criticized planning instruments of several cities in Sub-Saharan Africa, among them Kigali, as more responsive to the interest of real estate developers than those of the mostly poor urban residents.

Beyond ethical considerations, there is also a problem of the functioning of an urban system designed under such premises: the Kigali Master Plan (SURBANA International Consultants PTE Ltd., 2013) and, to a lesser degree (because it is less specific), the Kampala Physical Development Plan (ROM Transportation Engineering Ltd. et al., 2012) envision executing the strategy of compact development typical of planning practice in mature land markets (allow for a relatively high level of density in the center and constrain substantially development in the periphery, with denser inner suburbs and less dense outer suburbs and with certain mixed use subcenters to reduce congestion). This strategy relies for its execution on very high floor-to-area ratios in the city and very stringent constraints on building in the periphery. The latter has proven feasible in Sub-Saharan Africa (specifically in Kigali) but the former is very unlikely for at least two reasons. Firstly, developers have little incentive to substitute cheap labour with more expensive and productive capital (e.g. equipment such as cranes); this creates a technological barrier to the materialization of the supply of urban space foreseen in the regulation (and the demand for urban land still exists, which would then cause pressure for development in peripheral areas at lower density). Secondly, most urban agents are poor residents requiring housing; they do not have enough income to acquire and maintain living space in expensive structures such as multi-storey housing complexes. Since the planned supply is unlikely to materialize, this creates a problem of imbalance between supply and demand for an urban area. To date, efforts by Rwandan authorities to promote urban development have caused displacement of the urban poor (to formally built houses in peripheral locations, which enrich real estate developers through government subsidies while undermining local artisanal contractors from the informal sector; Nikuze et al., 2019) and the provision of financing for the high end segment of the real estate market to incentivize construction (including by a party-owned company, illustrating the political elite's stake on real estate development; Behuria and Goodfellow, 2019). Development potential in this high end segment of the market already shows signs of having reached its limit (Behuria and Goodfellow, 2019).

The consequences of this mismatch were explored using prospective scenarios of urban growth in chapter 7. The scenarios were designed to understand the differences on urban patterns introduced by the land use regulation instruments (the Kampala Physical Development Plan and the Kigali Master Plan). Since such prospective development depends critically on the level of densification, in turn a function of the city's overall population, two sets of scenarios for each city were created: for unplanned and for planned conditions; within each set, different scenarios consider high or low population growth and densification equal to or greater than the baseline. The resulting 2030 built-up land cover projections show the land use regulations of Kampala clearly promote a much more compact urban footprint. However, and unexpectedly, the land use plan of Kigali results in the dispersion of development to envisioned subcenters in the periphery of the current city. This is likely the result of an existing trend towards compaction, reflected in the unplanned scenarios, combined with a land use regulation instrument planned for a city with far greater population than what is to be expected for Kigali. In this sense, the projections also underscore the importance of population projections as a critical uncertainty but also the importance of context (Kigali by 2030 is still projected to present a smaller population than Kampala at present).

The impact of regulation on potential flooding was found to be limited for scenarios that did not incorporate a feedback between urban growth and flood. Urban development was found to be, ultimately, a function of land demand (population) with central locations generally developed at high density (built-up fractions generally over 0.50); at these levels of development, the runoff fraction of rainfall will be generally constant and high. Since no such feedback was expected to operate for Kigali (a relatively small city), under short run conditions land use regulation will not mitigate flooding because major flood problems will not have become apparent (i.e. such mitigation is unnecessary). As for Kampala, when incorporating the feedback between flooding and urban growth, two consequences follow: unplanned scenarios result in the development of most wetland areas (flat areas no longer constrained from development unless they recurrently flood) and the planned scenario, in addition to promoting compact development, also reflects the regulation's role in defending the wetland areas from development. When taken together, these projections result in a substantial reduction of built-up area exposed to flooding (of up to 25% under the hypothesis of high population growth).

Methodological and scientific contributions developed in the course of this dissertation can be divided into two main groups: firstly, several interesting results describing the case studies through the application of statistical techniques contributed to extend the state of the art in each of these fields and to increase knowledge about the specific cases (Kampala or Kigali). Secondly, the computational modeling and simulation tools – the main methodological development of this dissertation – represented an advance in several directions.

Structural equations modeling and the specifc ideas behind the model developed in chapter 3 (in particular, Shipley's contention that causal graphs should be acyclic and, in his view, feedbacks are in fact a causal chain incorrectly compressed in time) are not novel. However, their application to the analysis of landscape has been infrequent. The only precedent of such analysis in Sub-Saharan Africa found in the course of this dissertation was Odongo et al. (2014), who applied a similar analysis to water balances in Lake Naivasha. Many studies exist applying structural equations models to spatial data in the field of transportation (with records being locations or road segments); but hardly any were found using catchments as records in the field of hydrology (most applications analyze surveys used to determine vulnerability characteristics rather than the physical relations between units of the landscape). Furthermore, and most importantly, the main conclusion of this analysis (the detection of a weak but statistically significant constraint of flooding on urban growth in Kampala) provides important evidence to verify theoretical expectations and to understand the context in which they play out (particularly the interaction with population growth which proved to be so important).

The difference-in-differences model represents the first application of Kolak's (2007) causal framework to a spatial statistical model of urban growth. Previous models at "plot" (location) level had been developed for Oregon and Washington State in the United States (Dempsey and Plantinga, 2013; Kline et al., 2014) but none had, as yet, implemented fixed effects to control for spatial heterogeneity; the use of spatial lags, as a control, has been more common (although not in spatial statistical models of land cover change) but its interpretation has generally been lacking. It is important to stress that fully implementing the spatial causal framework (Kolak, 2017) changes the results relative to pooled models (the methodological equivalent of previous studies); specifically, the causal effects detected in models with fixed or random effects to control for spatial heterogeneity result in difference-in-differences estimators much smaller (1/5 to 1/10) than in pooled models, regardless of the inclusion of a spatial lag term.

The cellular automata model designed and implemented to model urban growth in Kampala and Kigali introduced a number of innovations, necessary to reflect the urban patterns and to integrate the results with the flood model.

Firstly, the cellular automata model conceives space as an array of cells each with an associated fraction of land cover (for built-up, vegetation, and soil). Continuous variables had been proposed for cellular automata modeling of urban systems by van Vliet et al. (2012), Yeh and Li (2001), and Li and Yeh (2000) to model the amount of activity in a location (e.g. the number of residents or of jobs in each cell) and the density; Li and Yeh (2000) in particular used so-called "grey cells" to incorporate the fraction of urban land into their model. Other than these three cases, however, the use of continuous variables in cellular automata modeling of land has been rare – and no precedent was found of models encoding land covers of the natural environment as well as of human activity. In the model developed (and reported in chapters 4, 5, and 7), such encoding of the dependent variable was required because it became directly the input of the flood model.

Secondly, the designed model, as most cellular automata models of urban growth, follows White (1998) in using ancillary information and a relatively standard set of spatial determinants including accessibility (Geographical Sciences Committee and others, 2014); however, and unlike most other applications, it explicitly reduces problems of correlation between the spatial determinants that embody the scaling relations which control urban expansion, namely accessibility and neighborhood effects; to do so, the first principal component of the neighborhood and accessibility factors was used as a determinant instead of all three maps simultaneously (this innovation was implemented when calibrating the model because the correlation caused problems with the Bayesian methods used to calibrate the model).

Thirdly, the model's intent of replicating the aggregate behavior of urban agents in a land market was implemented by using the traditional approach (a suitability map synthesizing the information used by urban agents to choose a location within the city) but it also explicitly accounted for the potential supply of urban land (i.e. the developer's possibilities to offer urban land in the desired locations). Thus, the model does not assume spatial equilibrium, in the sense that the market is capable of supplying demand wherever it is desired. Unlike the suitability map, which is linked to urban theory (Brueckner, 1987) by the information chosen to define the suitability index, the potential supply maps were either generated to replicate the observed randomness of Kampala and Kigali (in turn a consequence of urban agents mostly building their own houses) or to simulate the possibilities of regulation in supplying urban land – and it was this second capability which justified the explicit separation between supply and demand.

To calibrate the proposed cellular automata model (to determine the importance of each spatial factor and the size of the moding window that determines the neighborhood effect), Bayesian methods were applied. Specifically, a Markov Chain Monte Carlo approach was chosen. Only one other partial implementation of this method (Mustafa et al., 2017) was found and it used Markov Chain Monte Carlo only to calibrate the neighborhood effect (the weights of ancillary data were established by means of statistical analysis); more broadly, Bayesian sequential methods have also been used by Verstegen et al. (2014) to calibrate land models (the development of the Markov Chain Monte Carlo methodology drew heavily from this study, especially when defining how to compare the land patterns to externally derived data). No other precedents of applications to case studies were found in the literature in the development of this dissertation.

The use of land cover fractions as inputs for the urban growth model (and for the flood model) required a conceptual reflection on the appropriate methodology to produce the land cover maps. A first decision was taken to adopt mid-resolution Landsat satellite imagery as data to develop the land cover data models: because they are available for any extent potentially desired and because they have a long record, stretching back to the late 1970s. Spectral linear unmixing was chosen because of the possibility to produce land cover fractions at sub-pixel level (i.e. at an increased spatial resolution, as required by the flood model) by making use of the multi-spectral imagery; however, it must be noted that these methods were developed for hyper-spectral data: because the amount of possible fractions is limited by the available reflectance bands and because at least four fractions were required and only six bands were available (the thermal band was discarded), the implementation of the spectral mixing analysis relied on careful choice of endmember samples (the locations where the fractions existed in pure form). It was possible to derive land cover maps for both cities consistently in time for each city and the entire study period, even if the general methodological approach had to be adapted to each case. In particular, the Sub-Saharan African context implied greater presence of bare soil (relative to urban areas of the industrialized north, for which the methods were originally developed); this is a problem because of the high albedo of some soils, which can be easily confused with buildings. In Kigali, the use of the Global Human Settlements Layer of the Joint Research Center (Pesaresi et al., 2016) to interpret the fraction maps that were created contributed to mitigate this problem.

The integrated modeling tool used to simulate the feedback between urban growth and flooding represents a rare spatially explicit implementation of two way causality. Studies analyzing the effect of flooding on land patterns are relatively uncommon, as was discussed in chapter 3. No precedent was found of a spatially explicit recursive formulation (i.e. one updating land patterns and flooding at each period simulated). The scenarios developed with this tool were few and purposefully chosen to distinguish the consequences of implementing the land use plan. However, the modeling tool can be used as a methodological framework to explore uncertainties associated to a number of (physical) factors, in particular:

- A single succession of rainfall events, randomly ordered, was used for the three scenarios. The order of these events, i.e. how rainfall was simulated to occur in time, potentially has large effects on land cover trends. In particular, it is important whether large rainfall events are simulated to occur in a short time span or spread out over the entire simulation period (the latter was the case for the results of chapter 7) and also whether the larger events occur early in the simulation or later. Rainfall uncertainty in time (at the scale of the event occurrence) can and should be explored using the integrated model as a framework.

- In Kampala (and in Kigali) rainfall does not fall uniformly over the entire city at the same time. Large differences may occur between diverse locations. *OpenLISEM* is capable of handling events inputed as rainfall maps with each pixel the rainfall intensity at that location differing from other locations (the implementation in chapter 7 assumes, as most hydrological applications, a uniform rate falling over the entire city). Rainfall uncertainty in space is a second area in need of work within the capability of the integrated model.

- The rainfall events simulated correspond to design storms generated by application of the intensity-duration-frequency methodology, which are symmetric in time. Rainfall events measured for Kampala (Sliuzas et al., 2013) do not follow this simple shape (see also the analysis by Turkington, 2016 on an area with richer data). Rainfall uncertainty in time (at the scale of the event intensity) may become a critical factor and should be explored.

- While no measurement of yearly maximum rainfall events at detailed (10-15 min) intervals exist for Kampala, it is possible to reconstruct a series of events under the intensity-duration-frequency assumption for the past (2001-2016) that approximate the maximum rainfall event for each year. With this information, it should be possible to apply the Monte Carlo calibration methodology, described in chapter 5, to derived a flooding factor and a net wetland factor (excluding the flood's effect).

- The main obstacle to the implementation of a calibration process with a flood factor is the duration of the simulation (a full 14 year simulation for Kampala was achieved in a high performance laptop in approximately 15 hours of computational time; in contrast, each simulation of the calibration analysis was completed in less than 5 minutes). In addition to improving the calibrated model by elucidating the net contribution of flooding and non-flooding wetland effects, work is also required to increase the computational speed of the integrated tool.
- More generally, Umer et al. (2019) performed a sensitivity analysis; some of the factors they identified may also need to be explored as sources of potential spatial or temporal uncertainty (soil characteristics proved to have limited impact; rainfall patterns and events, however, may be key from an uncertainty perspective).

Conclusions for land use planning design and implementation in Kampala and Kigali were developed in the course of this dissertation. A first important gap filled by the outcomes of this study is the modeling tools: an integrated urban growth and flood model for Kampala, a cellular automata-based urban growth model for Kigali, and the information compiled to develop them (particularly the maps of land cover fractions for three periods of Kigali and four periods of Kampala). This information and the tools can be used to test land use regulation, as was demonstrated in chapter 7 and in Pérez-Molina et al. (2016).

The second important conclusion relates to the design of land use planning instruments in Kampala and Kigali. The Kigali Master Plan clearly envisions a city based on multi-storey buildings for most of its dense areas. This strategy is very likely not feasible, given the income levels of most urban residents. Moreover, as was demonstrated through simulation, there is a danger in designing a city with so much development potential: that areas planned to be subcenters of a large city become isolated high density enclaves and that this urban form produces a sprawling pattern highly reliant on motorized transportation. A similar strategy was sketched for Kampala in its physical development plan – and although Kampala is a larger city with some multi-storey buildings already in existence, to excessively rely on this type of buildings is unlikely to achieve positive results in terms of urban form.

The third finding of this dissertation relates to the role of land use planning to mitigate flood risk. The simulations of prospective land cover patterns suggest relatively high fractions of built-up land cover as both cities expand, under planned and unplanned conditions. This means it will be very difficult to reduce runoff. Infrastructure could contribute to increase infiltration or to slow the flow of water over the landscape, but should be expected to cause limited impacts. Therefore, the best strategy to reduce flood risk is to protect natural wetlands and floodplains from urban development, a strategy already being implemented successfully in Kigali (although this city's authorities should also be forewarned of the expected increase in pressure to occupy such areas as the city becomes larger).

Finally, infrastructure (particularly drainage infrastructure) is a critical factor in mitigating flood risk. In addition to relatively large projects to improve the primary and secondary drainage systems, it is also important for both Kampala and Kigali to gradually create a network of local storm drainage (for every street in each of the cities). While this task is daunting, it will likely require more patience and systematic investment (in construction and in maintenance) rather than quick and effective one time decisions. It is also not a highly visible investment. It is thus a test to the institutional robustness of both municipalities (of Kampala and of Kigali) to sustain such an effort. However, it is also unavoidable if the problems caused by recurrent flooding are to be successfully tackled: because the overwhelming majority of floods that affect the functioning of these cities (and this is especially true of Kamapala) are not the mid-to-large events with return periods of 1:10 years. Rather, they are the much smaller and frequent torrential rainfalls that occur often during the wet season. The infrastructure solutions have, to date, concentrated on the larger systems but these smaller, recurrent impacts are likely a worse problem in terms of urban functionality (the scale of the solution has been metropolitan when the problem is actually at neighborhood, or perhaps better, at block level). Land use regulation can contribute to this program of investment through requirements for developers to build the storm drainage at their own cost when creating urban land through the lot subdivision process.

A coda: In addition to synthesizing what this dissertation became, in the process of exploring urban growth, flooding, and spatial planning in Kampala and Kigali, it is also of value to record how this dissertation was originally planned and why and how this plan was changed. Originally, this project was organized to compare two sub-catchments, each within one metropolitan region: Upper Lubigi in Kampala and Rwampara in Kigal. These areas were chosen because of the presence of informal settlements in the lower part of the sub-catchment and the potential this arrangement had of representing the prototypical recurrent flooding in the metropolitan region, as well as because sufficient area existed for urban expansion to take place.

However, when analyzing urban growth patterns, it became clear that urban expansion is a phenomenon occurring at metropolitan scale: the range of spatial variation that needs to be accounted for can only be quantitatively explored if the entire metropolitan region is analyzed, else there is not enough variance in the data. Additionally, as has been repeatedly argued in the course of this dissertation, the overall population of the city is a critical characteristic; but this is only apparent at metropolitan level. To deploy prospective simulation with urban growth models, then, a limitation identified is the spatial extent: these methods are appropriate for the full extent of a relatively large urban region; if applied to a part of the region, they should be carefully examined for pertinence and efficiency.

The challenges of Kampala and Kigali in their uncertain road to development are enormous. Land use planning is one of many possible collective actions towards the goal. Whether the citizens of Kampala and of Kigali opt to include it, or not, in their development strategy, one can but wish them well in their endeavour and hope the knowledge developed in this dissertation becomes an aid to build a better future for themselves and their cities.

Bibliography

Abebe, G. A. (2013). Quantifying urban growth pattern in developing countries using remote sensing and spatial metrics: A case study in Kampala, Uganda. Master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

African Planning Association (2014). *The State of Planning in Africa. An overview*. APA and UN-Habitat, Nairobi, Kenya.

Alberti, M., Asbjornsen, H., Baker, L. A., Brozovic, N., Drinkwater, L. E., Drzyzga, S. A., Jantz, C. A., Fragoso, J., Holland, D. S., Kohler, T. T. A., et al. (2011). Research on coupled human and natural systems (CHANS): approach, challenges, and strategies. *The Bulletin of the Ecological Society of America*, 92(2):218–228.

Anderson, J. R. (1976). *A Land Use and Land Cover Classification System for Use with Remote Sensor Data*, volume 964 of *Geological Survey Professional Paper*. US Government Printing Office.

Anselin, L. (1999). Spatial Econometrics. Working Paper. Bruton Center, School of Social Sciences, University of Texas at Dallas.

Anthony, J. (2017). Housing price effects of growth regulations: a concise taxonomy. *International Journal of Housing Policy*, 17(4):569–590.

Ashenfelter, O. and Card, D. (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *The Review of Economics and Statistics*, 67(4):648–660.

Baartman, J. E., Jetten, V. G., Ritsema, C. J., and Vente, J. (2012). Exploring effects of rainfall intensity and duration on soil erosion at the catchment scale using openLISEM: Prado catchment, SE Spain. *Hydrological Processes*, 26(7):1034–1049.

Baptista, R. (2003). Productivity and the Density of Local Clusters. In Bröcker, J., Dohse, D., and Soltwedel, R., editors, *Innovation Clusters and Interregional Competition*, Advances in Spatial Science, pages 163–181. Springer, Berlin & Heidelberg.

Bathrellos, G. D., Skilodimou, H. D., Chousianitis, K., Youssef, A. M., and Pradhan, B. (2017). Suitability estimation for urban development using multi-hazard assessment map. *Science of the Total Environment*, 575:119–134.

Batty, M. (2009). Cities as Complex Systems: Scaling, Interaction, Networks, Dynamics and Urban Morphologies. In Meyers, R. A., editor, *Encyclopedia of Complexity and Systems Science*, pages 1041–1071. Springer, New York, NY.

Baum-Snow, N. and Ferreira, F. (2015). Causal inference in urban and regional economics. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5, pages 3–68. Elsevier, Amsterdam.

Behuria, P. and Goodfellow, T. (2019). Leapfrogging Manufacturing? Rwanda's Attempt to Build a Services-Led 'Developmental State'. *The European Journal of Development Research*, pages 1–23. Accepted for publication.

Bhatta, B. (2010). *Analysis of Urban Growth and Sprawl from Remote Sensing Data*. Advances in Geographic Information Science. Springer Science & Business Media, Heidelberg.

Binder, C., Hinkel, J., Bots, P., and Pahl-Wostl, C. (2013). Comparison of Frameworks for Analyzing Social-ecological Systems. *Ecology and Society*, 18(4).

Bizimana, J. P. and Schilling, M. (2010). Geo-Information Technology for Infrastructural Flood Risk Analysis in Unplanned Settlements: a case study of informal settlement flood risk in the Nyabugogo flood plain, Kigali City, Rwanda. In Showalter, P. S. and Lu, Y., editors, *Geospatial Techniques in Urban Hazard and Disaster Analysis*, volume 2 of *Geotechnologies and the Environment*, pages 99–124. Springer, Dordrecht.

Brown, D. G., Page, S., Riolo, R., Zellner, M., and Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science*, 19(2):153–174.

Brueckner, J. K. (1983). The economics of urban yard space: An "implicit-market" model for housing attributes. *Journal of Urban Economics*, 13(2):216–234.

Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the Muth-Mills model. In Mills, E. S., editor, *Handbook of Regional and Urban Economics*, volume 2, chapter 20, pages 821–845. North Holland, Amsterdam.

Brueckner, J. K. (2000). Urban Sprawl: Diagnosis and Remedies. *International Regional Science Review*, 23(2):160–171.

Brueckner, J. K. and Sridhar, K. S. (2012). Measuring welfare gains from relaxation of land-use restrictions: The case of India's building-height limits. *Regional Science and Urban Economics*, 42(6):1061–1067.

Cai, H., Wang, Z., and Zhang, Q. (2017). To build above the limit? Implementation of land use regulations in urban China. *Journal of Urban Economics*, 98:223–233.

Campanella, T. J. (2006). Urban resilience and the recovery of New Orleans. *Journal of the American Planning Association*, 72(2):141–146.

Chaudhuri, G. and Clarke, K. (2013). The SLEUTH land use change model: A review. *Environmental Resources Research*, 1(1):88–105.

Chaussé, P. (2010). Computing Generalized Method of Moments and Generalized Empirical Likelihood with R. *Journal of Statistical Software*, 34(11):1–35.

Chen, X., Wang, D., Tian, F., and Sivapalan, M. (2016). From channelization to restoration: Sociohydrologic modeling with changing community preferences in the Kissimmee River Basin, Florida. *Water Resources Research*, 52(2):1227–1244.

Chereni, S. (2016). Determinants of in-situ flood damage mitigation in Bwaise 3, Kampala Uganda. Master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, 49(4):327–335.

Chogyal, J. (2012). The Current and Future Flood Situation, Bwaise III, Kampala, Uganda, Africa. Master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

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.

Ciavola, S. J., Jantz, C. A., Reilly, J., and Moglen, G. E. (2014). Forecast changes in runoff quality and quantity from urbanization in the DelMarVa Peninsula. *Journal of Hydrologic Engineering*, 19(1):1–9.

Clarke, K. C., Hoppen, S., and Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B: Planning and Design*, 24(2):247–261.

Cobbinah, P. B., Erdiaw-Kwasie, M. O., and Amoateng, P. (2015). Africa's urbanisation: Implications for sustainable development. *Cities*, 47:62–72.

Congdon, P. (2006). *Bayesian Statistical Modelling*. John Wiley & Sons, Chichester, UK, 2nd edition.

Cowen, D. J. and Jensen, J. R. (1998). Extraction and Modeling of Urban Attributes Using Remote Sensing Technology. In Liverman, D., Moran, E. F., Rindfuss, R. R., and Stern, P. C., editors, *People and Pixels. Linking Remote Sensing and Social Science*, pages 64–188. National Academy Press, Washington DC.

De Jong, S. M. and Jetten, V. G. (2007). Estimating spatial patterns of rainfall interception from remotely sensed vegetation indices and spectral mixture analysis. *International Journal of Geographical Information Science*, 21(5):529–545.

De Roo, A. P. J., Wesseling, C. G., and Ritsema, C. J. (1996). LISEM: A single-event physically based hydrological and soil erosion model for drainage basins. I: Theory, input and output. *Hydrological Processes*, 10(8):1107–1117.

Delestre, O., Cordier, S., Darboux, F., Du, M., James, F., Laguerre, C., Lucas, C., and Planchon, O. (2014). FullSWOF: A software for overland flow simulation. In Gourbesville, P., Cunge, J., and Caignaert, G., editors, *Advances in Hydroinformatics. SIMHYDRO 2012-New Frontiers of Simulation*, pages 221–231. Springer, Singapore.

Dempsey, J. A. and Plantinga, A. J. (2013). How well do urban growth boundaries contain development? Results for Oregon using a difference-in-difference estimator. *Regional Science and Urban Economics*, 43(6):996–1007.

Di Baldassarre, G., Viglione, A., Carr, G., Kuil, L., Salinas, J., and Blöschl, G. (2013). Socio-hydrology: conceptualising human-flood interactions. *Hydrology and Earth System Sciences*, 17(8):3295–3303.

Dodman, D., Leck, H., Rusca, M., and Colenbrander, S. (2017). African urbanisation and urbanism: Implications for risk accumulation and reduction. *International Journal of Disaster Risk Reduction*, 26:7–15.

Douglas, I., Alam, K., Maghenda, M., Mcdonnell, Y., Mclean, L., and Campbell, J. (2008). Unjust waters: climate change, flooding and the urban poor in Africa. *Environment and Urbanization*, 20(1):187–205.

Du, Q., Wasson, L., and King, R. (2005). Unsupervised linear unmixing for change detection in multitemporal airborne hyperspectral imagery. In *International Workshop on the Analysis of Multi-Temporal Remote Sensing Images, 2005*, pages 136–140. IEEE.

Edmonds, B. and Moss, S. (2005). From KISS to KIDS–An 'Anti-simplistic' Modelling Approach. In Davidsson, P., Logan, B., and Takadama, K., editors, *Multi-Agent and Multi-Agent-Based Simulation*, number 3415 in Lecture Notes in Artificial Intelligence, pages 130–144. Springer-Verlag, Berlin, Germany.

Elga, S., Jan, B., and Okke, B. (2015). Hydrological modelling of urbanized catchments: a review and future directions. *Journal of Hydrology*, 529:62–81.

Fiddes, D. and Forsgate, J. A. (1974). The prediction of storm rainfall in east africa. Technical Report TRRL Laboratory Report 623., Transport and Road Research Laboratory, Crowthorne, UK.

Filatova, T., Parker, D. C., and van der Veen, A. (2011). The implications of skewed risk perception for a Dutch coastal land market: insights from an agent-based computational economics model. *Agricultural and Resource Economics Review*, 40(3):405–423.

Frame, D. E. (1998). Housing, Natural Hazards, and Insurance. *Journal of Urban Economics*, 44(1):93–109.

Fura, G. D. (2013). Analysing and modelling urban land cover change for run-off modelling in Kampala, Uganda. Master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472.

Geographical Sciences Committee and others (2014). *Advancing Land Change Modeling: Opportunities and Research Requirements*. National Academies Press, Washington DC.

Gilks, W. R. (1996). Full conditional distributions. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 75–88. Chapman & Hall/CRC, Boca Raton, FL.

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994). A language and program for complex Bayesian modelling. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 43(1):169–177.

Glaeser, E. L. (2008). *Cities, Agglomeration, and Spatial Equilibrium*. The Lindahl Lectures. Oxford University Press, Oxford, UK.

Glaeser, E. L. and Ward, B. A. (2009). The causes and consequences of land use regulation: Evidence from Greater Boston. *Journal of Urban Economics*, 65(3):265–278.

Glaeser, E. L. and Xiong, W. (2017). Urban productivity in the developing world. *Oxford Review of Economic Policy*, 33(3):373–404.

Gollin, D., Jedwab, R., and Vollrath, D. (2016). Urbanization with and without Industrialization. *Journal of Economic Growth*, 21(1):35–70.

Goodfellow, T. (2012). *State effectiveness and the politics of urban development in East Africa. A puzzle of two cities, 2000-2010*. PhD thesis, London School of Economics and Political Science, London, UK.

Goodfellow, T. (2013a). Planning and development regulation amid rapid urban growth: explaining divergent trajectories in Africa. *Geoforum*, 48:83–93.

Goodfellow, T. (2013b). Urban planning in Africa and the politics of implementation: contrasting patterns of state intervention in Kampala and Kigali. In Obrist, B., Arlt, V., and Macamo, E., editors, *Living the City in Africa: Processes of Invention and Intervention*, Schweizerische Afrikastudien - Etudes africaines suisses, pages 45–62. LIT Verlag, Zurich, Switzerland.

Goodfellow, T. (2017). Taxing property in a neo-developmental state: The politics of urban land value capture in Rwanda and Ethiopia. *African Affairs*, 116(465):549–572.

Goodfellow, T. (2018). Seeing Political Settlements through the City: A Framework for Comparative Analysis of Urban Transformation. *Development and Change*, 49(1):199–222.

Goodfellow, T. and Smith, A. (2013). From urban catastrophe to 'model' city? politics, security and development in post-conflict kigali. *Urban studies*, 50(15):3185–3202.

Greene, W. H. (2003). *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 5th edition.

Güneralp, B., Lwasa, S., Masundire, H., Parnell, S., and Seto, K. C. (2017). Urbanization in Africa: challenges and opportunities for conservation. *Environmental Research Letters*, 13(1):015002.

Habonimana, H. V. (2014). Integrated flood modeling in Lubigi catchment Kampala. Master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

Habonimana, H. V., Bizimana, J. P., Uwayezu, E., Tuyishimire, J., and Mugisha, J. (2015). Integrated flood modeling for flood hazard assessment in Kigali City, Rwanda. In *Proceedings of GeoTech Rwanda 2015*, Kigali, Rwanda. ITC and University of Rwanda.

Hagen-Zanker, A. and Martens, P. (2008). Map comparison methods for comprehensive assessment of geosimulation models. In Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., and Gavrilova, M. L., editors, *Computational Science and Its Applications–ICCSA 2008*, pages 194–209. Springer, Berlin & Heidelberg.

Hall, J., Arheimer, B., Borga, M., Brázdil, R., Claps, P., Kiss, A., Kjeldsen, T., Kriauciuniene, J., Kundzewicz, Z., Lang, M., et al. (2014). Understanding flood regime changes in Europe: A state of the art assessment. *Hydrology and Earth System Sciences*, 18(7):2735–2772.

Hansen, H. S. (2014). Modelling the future coastal zone urban development as implied by the IPCC SRES and assessing the impact from sea level rise. *Landscape and Urban Planning*, 98(3):141–149.

Hartig, F., Minunno, F., and Paul, S. (2018). *BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics*. R package version 0.1.5.

Hessel, R., Jetten, V., Liu, B., Zhang, Y., and Stolte, J. (2003). Calibration of the LISEM model for a small Loess Plateau catchment. *Catena*, 54(1):235–254.

Hirschman, A. O. (1985). Against Parsimony: Three Easy Ways of Complicating some Categories of Economic Discourse. *Economics & Philosophy*, 1(1):7–21.

Hitchcock, D. B. (2003). A History of the Metropolis-Hastings Algorithm. *The American Statistician*, 57(4):254–257.

Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Springer, Dordrecht, 2nd edition.

Hou, H., Estoque, R. C., and Murayama, Y. (2016). Spatiotemporal analysis of urban growth in three African capital cities: A grid-cell-based analysis using remote sensing data. *Journal of African Earth Sciences*, 123:381–391.

Huong, H. T. L. and Pathirana, A. (2013). Urbanization and climate change impacts on future urban flooding in Can Tho city, Vietnam. *Hydrology and Earth System Sciences*, 17(1):379–394.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

International Research Institute for Climate and Society, Columbia University (2018). Iri/ldeo climate data library. http://iridl.ldeo.columbia.edu/. Accessed 2018-10-18.

Isunju, J. B., Orach, C. G., and Kemp, J. (2016). Hazards and vulnerabilities among informal wetland communities in Kampala, Uganda. *Environment and Urbanization*, 28(1):275–293.

Jackson, K. (2016). Do land use regulations stifle residential development? Evidence from California cities. *Journal of Urban Economics*, 91:45–56.

Jetten, V. (2018). *OpenLISEM Multi-Hazard Land Surface Process Model. Documentation & User Manual*. Faculty of Geo-Information Science and Earth Observation, University of Twente., Enschede, The Netherlands.

Jetten, V. G. (2014). openLISEM – A spatial model for runoff, floods and erosion. The Rasterscript. `http://blogs.itc.nl/lisem/running-lisem/the-rasterscript/`. Online; accessed 26 Jan. 2016.

Jetten, V. G. (2016). Caribbean Handbook on Risk Information Management (CHaRIM) - Saint Lucia National Flood Hazard Map Methodology and Validation Report. ACP-EU National Disaster Risk Reduction Programme 2016. pp 43.

Jørgensen, S. E. and Bendoricchio, G. (2001). *Fundamentals of Ecological Modelling*, volume 21 of *Developments in Environmental Modelling*. Elsevier, Amsterdam, The Netherlands, 3rd edition.

Kahn, M. E., Vaughn, R., and Zasloff, J. (2010). The housing market effects of discrete land use regulations: Evidence from the California coastal boundary zone. *Journal of Housing Economics*, 19(4):269–279.

Karssenberg, D., Schmitz, O., Salamon, P., de Jong, K., and Bierkens, M. F. (2010). A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling & Software*, 25(4):489–502.

Khan, D. M., Veerbeek, W., Chen, A. S., Hammond, M. J., Islam, F., Pervin, I., Djordjević, S., and Butler, D. (2018). Back to the future: assessing the damage of 2004 Dhaka Flood in the 2050 urban environment. *Journal of Flood Risk Management*, 11:S43–S54.

Kim, H. and Millsap, R. (2014). Using the Bollen-Stine Bootstrapping Method for Evaluating Approximate Fit Indices. *Multivariate Behavioral Research*, 49(6):581–596.

Kline, J. D., Thiers, P., Ozawa, C. P., Yeakley, J. A., and Gordon, S. N. (2014). How well has land-use planning worked under different governance regimes? A case study in the Portland, OR-Vancouver, WA metropolitan area, USA. *Landscape and Urban Planning*, 131:51–63.

Kolak, M. A. (2017). *Policy and Place: A Spatial Data Science Framework for Research and Decision-Making*. PhD thesis, Arizona State University, Tempe, AZ, United States.

Kuang, W., Chi, W., Lu, D., and Dou, Y. (2014). A comparative analysis of megacity expansions in China and the U.S.: Patterns, rates and driving forces. *Landscape and Urban Planning*, 132:212–135.

Kumar, D. S., Arya, D., and Vojinovic, Z. (2013). Modeling of urban growth dynamics and its impact on surface runoff characteristics. *Computers, Environment and Urban Systems*, 41:124–135.

Kutilek, M. and Nielsen, D. R. (1994). *Soil Hydrology*. Catena Verlag, Cremlingen, Germany.

Lall, S. V. (2017). Renewing expectations about Africa's cities. *Oxford Review of Economic Policy*, 33(3):521–539.

Lambin, E. F., Turner, B. L., Geist, H. J., Agbola, S. B., Angelsen, A., Bruce, J. W., Coomes, O. T., Dirzo, R., Fischer, G., Folke, C., George, P. S., Homewood, K., Imbernon, J., Leemans, R., Li, X., Moran, E. F., Mortimore, M., Ramakrishnan, P. S., Richards, J. F., Skånes, H., Steffen, W., Stone, G. D., Svedin, U., Veldkamp, T. A., Vogel, C., and Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change*, 11(4):261–269.

Lehnert, L. W., Meyer, H., and Bendix, J. (2016). *hsdar: Manage, analyse and simulate hyperspectral data in R*. R package version 0.5.1.

Li, X. and Yeh, A. G.-O. (2000). Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science*, 14(2):131–152.

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, Zhiyun 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.

Love, T. B., Kumar, V., Xie, P., and Thiaw, W. (2004). A 20-year daily Africa precipitation climatology using satellite and gauge data. In *14th Conference on Applied Climatology*. American Meteorological Society.

Lwasa, S. (2010). Adapting urban areas in Africa to climate change: the case of Kampala. *Current Opinion in Environmental Sustainability*, 2(3):166–171.

Mabikke, S. B. (2016). Historical continuum of land rights in uganda: A review of land tenure systems and approaches for improving tenure security. *Journal of Land and Rural Studies*, 4(2):153–171.

Magliocca, N. R., Rudel, T. K., Verburg, P. H., McConnell, W. J., Mertz, O., Gerstner, K., Heinimann, A., and Ellis, E. C. (2015). Synthesis in land change science: methodological patterns, challenges, and guidelines. *Regional Environmental Change*, 15(2):211–226.

Mamadouh, V. (1999). Grid-group cultural theory: an introduction. *GeoJournal*, 47(3):395–409.

McColl, C. and Aggett, G. (2007). Land-use forecasting and hydrologic model integration for improved land-use decision support. *Journal of Environmental Management*, 84(4):494–512.

McHarg, I. L. (1969). *Design with Nature*. American Museum of Natural History, New York.

Meadows, D. H., Meadows, D. H., Randers, J., and Behrens III, W. W. (1972). *The Limits to Growth. A Report for the Club of Rome's Project on the Predicament of Mankind.* Universe Books, New York, NY.

Meyer, W. B. and Turner, B. L. (1990). Regional studies of transformation. Editorial Introduction. In Turner, B. L., Clark, W. C., Kates, R. W., Richards, J. F., Mathews, J. T., and Meyer, W. B., editors, *The Earth As Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*, pages 369–472. Cambridge University Press, Cambdrige, UK.

Millo, G. and Piras, G. (2012). splm: Spatial Panel Data Models in R. *Journal of Statistical Software*, 47(1):1–38.

Mohnda, A. (2013). Evaluating Flash Flood Risk Reduction Strategies in Built-up Environment in Kampala. Master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

Mustafa, A., Nishida, G., Saadi, I., Cools, M., and Teller, J. (2017). A Markov Chain Monte Carlo Cellular Automata Model to Simulate Urban Growth. In Rückemann, C.-P., Doytsher, Y., and Braz, F. J., editors, *GEO-Processing 2017 proceedings, The Ninth International Conference on Advanced Geographic Information Systems, Applications, and Services*. International Academy, Research, and Industry Association (IARIA), Nice, France.

Nduwayezu, G., Sliuzas, R., and Kuffer, M. (2016). Modeling urban growth in Kigali city Rwanda. *Rwanda Journal, Series D*, 1(Special Edition 2):31pp.

Nigussie, T. A. and Altunkaynak, A. (2016). Assessing the hydrological response of Ayamama watershed from urbanization predicted under various landuse policy scenarios. *Water Resources Management*, 30(10):3427–3441.

Nikuze, A., Sliuzas, R., Flacke, J., and van Maarseveen, M. (2019). Livelihood impacts of displacement and resettlement on informal households-a case study from kigali, rwanda. *Habitat International*, 86:38–47.

Odongo, V. O., Mulatu, D. W., Muthoni, F. K., Van Oel, P. R., Meins, F. M., Van der Tol, C., Skidmore, A. K., Groen, T. A., Becht, R., Onyando, J. O., and van der Veen, A. (2014). Coupling socio-economic factors and eco-hydrological processes using a cascade-modeling approach. *Journal of Hydrology*, 518:49–59.

OZ Architecture, Shapira-Hellerman Planners, EDAW, Tetra Tech, ERA, and Engineers Without Borders (2007). *Kigali Conceptual Master Plan*. City of Kigali, Kigali, Rwanda.

Pande, S. and Sivapalan, M. (2017). Progress in socio-hydrology: a meta-analysis of challenges and opportunities. *WIREs Water*, 4(4):e1193.

Pebesma, E. J. and Wesseling, C. G. (1998). Gstat: a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, 24(1):17–31.

Pérez-Molina, E. (2014). Modeling urban growth and flooding interactions with cellular automata in Kampala, Uganda. Master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.

Pérez-Molina, E., Sliuzas, R., Flacke, J., and Jetten, V. (2017). Developing a cellular automata model of urban growth to inform spatial policy for flood mitigation: A case study in Kampala, Uganda. *Computers, Environment and Urban Systems*, 65:53–65.

Pérez-Molina, E., Sliuzas, R., Jetten, V., and van Maarseveen, M. (2019a). Structural equations modeling of hydrological impacts on urban land cover in Kampala, 2001-2016. Manuscript submitted for publication.

Pérez-Molina, E., Sliuzas, R., van Maarseveen, M., and Jetten, V. (2019b). A cellular automata model of urban growth for Sub-Saharan Africa: calibration and validation with a Markov chain Monte Carlo method. Manuscript submitted for publication.

Pérez-Molina, E., Sliuzas, R., van Maarseveen, M. F. A. M., and Jetten, V. (2016). Urban flood management under rapid growth in Kigali, Rwanda: developing perspectives on land use planning and the future by exploring the case of the Rwampara wetland. In Schmitt, H. C., Danielzyk, R., Greiving, S., Gruehn, D., Thinh, N. X., and Warner, B., editors, *Raummuster. Struktur–Dynamik–Planung*, volume 147 of *Blaue Reihe. Dortmunder Beiträge zur Raumplanung*, pages 351–365. Klartext, Essen, Germany.

Pesaresi, M., Syrris, V., and Julea, A. (2016). A New Method for Earth Observation Data Analytics Based on Symbolic Machine Learning. *Remote Sensing*, 8(5):399.

Pitt, M. A., Myung, I. J., and Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, 109(3):472.

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11.

Poelmans, L., Rompaey, A. V., Ntegeka, V., and Willems, P. (2011). The relative impact of climate change and urban expansion on peak flows: a case study in central Belgium. *Hydrological Processes*, 25(18):2846–2858.

Pontius Jr, R. G. (2002). Statistical methods to partition effects of quantity and location during comparison of categorical maps at multiple resolutions. *Photogrammetric Engineering and Remote Sensing*, 68(10):1041–1050.

Potts, D. (2009). The slowing of sub-Saharan Africa's urbanization: evidence and implications for urban livelihoods. *Environment and Urbanization*, 21(1):253–259.

Pulver, S., Ulibarri, N., Sobocinski, K. L., Alexander, S. M., Johnson, M. L., McCord, P. F., and Dell'Angelo, J. (2018). Frontiers in socio-environmental research: components, connections, scale, and context. *Ecology and Society*, 23(3).

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rasmussen, R. and Hamilton, G. (2012). An approximate Bayesian computation approach for estimating parameters of complex environmental processes in a cellular automata. *Environmental Modelling & Software*, 29(1):1–10.

Refsgaard, J. C. and Henriksen, H. J. (2004). Modelling guidelines–terminology and guiding principles. *Advances in Water Resources*, 27(1):71–82.

Ridd, M. K. (1995). Exploring a V-I-S (vegetation-impervious surface-soil model for urban ecosystem analysis through remote sensing: comparative anatomy for cities. *International Journal of Remote Sensing*, 6(12):2165–2185.

Rodrigues, H. and Soares-Filho, B. (2018). A short presentation of dinamica ego. In Camacho Olmedo, M. T., Paegelow, M., Mas, J.-F., and Escobar, F., editors, *Geomatic Approaches for Modeling Land Change Scenarios*, pages 493–498. Springer International Publishing, Cham.

ROM Transportation Engineering Ltd., Shapira-Hellerman Planners, Larry Aberman & Assiociates, Tzamir Architects and Planners Ltd., and Ofek Aerial Photography Ltd. (2012). *Kampala Physical Development Plan*. Kampala Capital City Authority, Kampala, Uganda.

Rosenblueth, A. and Wiener, N. (1945). The Role of Models in Science. *Philosophy of Science*, 12(4):316–321.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2):1–36.

Rykiel, E. J. (1996). Testing ecological models: the meaning of validation. *Ecological Modelling*, 90(3):229–244.

Sakieh, Y., Salmanmahiny, A., and Mirkarimi, S. H. (2017). Tailoring a non-path-dependent model for environmental risk management and polycentric urban land-use planning. *Environmental Monitoring and Assessment*, 189(2):91.

Sánchez-Moreno, J. F., Jetten, V. G., Mannaerts, C. M., and Pina Tavares, J. (2014). Selecting best mapping strategies for storm runoff modeling in a mountainous semi-arid area. *Earth Surface Processes and Landforms*, 39(8):1030–1048.

Santé, I., García, A. M., Miranda, D., and Crecente, R. (2010). Cellular automata models for the simulation of real-world urban processes: A review and analysis. *Landscape and Urban Planning*, 96(2):108–122.

Savalei, V. (2014). Understanding Robust Corrections in Structural Equation Modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(1):149–160.

Seasholtz, M. B. and Kowalski, B. (1993). The parsimony principle applied to multivariate calibration. *Analytica Chimica Acta*, 277(2):165–177.

Sene, K. (2010). *Hydrometeorology. Forecasting and Applications*. Springer, Heidelberg, 2nd edition edition.

Severen, C. and Plantinga, A. J. (2018). Land-use regulations, property values, and rents: Decomposing the effects of the California Coastal Act. *Journal of Urban Economics*, 107:65–78.

Shearer, A. W., Mouat, D. A., Bassett, S. D., Binford, M. W., Johnson, C. W., Saarinen, J. A., Gertler, A. W., and Kahyaoğlu-Koračin, J. (2009). *Land Use Scenarios: Environmental Consequences of Development*. Integrative Studies in Water Management and Land Development. CRC Press, Boca Raton, FL.

Shipley, B. (2016). *Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference with R*. Cambridge University Press, Cambridge, UK, 2nd edition.

Shipley, B. and Lechowicz, M. J. (2000). The functional co-ordination of leaf morphology, nitrogen concentration, and gas exchange in 40 wetland species. *Ecoscience*, 7(2):183–194.

Signé, L. (2018). Why has Africa's fast economic growth left more poor people behind and how do we fix it? In Coulibaly, B. S., editor, *Foresight Africa. Top Priorities for the Continent in 2018*, chapter 3, pages 44–53. Brookings Institution, Washington DC.

Silva, C. N. (2012). Urban Planning in Sub-Saharan Africa: An Overview. *Cities*, 29(3):155–157.

Sivapalan, M., Savenije, H. H., and Blöschl, G. (2012). Socio-hydrology: A new science of people and water. *Hydrological Processes*, 26(8):1270–1276.

Sliuzas, R. (2004). *Managing Informal Settlements. A Study Using Geo-Information in Dar es Salaam, Tanzania.* PhD thesis, Utrecht University and ITC, Enschede, The Netherlands.

Sliuzas, R., Flacke, J., and Jetten, V. (2013). Modelling urbanization and flooding in Kampala, Uganda. In *Proceedings of the 14th N-AERUS / GISDECO conference*, Enschede, The Netherlands. N-AERUS / GISDECO.

Small, C. (2001). Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing*, 22(7):1305–1334.

Smith, K. and Ward, R. (1998). *Floods. Physical Processes and Human Impacts.* Wiley, Chichester, UK.

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.

Somodi, I., Virágh, K., and Miklós, I. (2011). A Bayesian MCMC approach to reconstruct spatial vegetation dynamics from sparse vegetation maps. *Landscape Ecology*, 26(6):805–822.

Southall, A. W. and Gutkind, P. C. W. (1957). *Townsmen in the Making. Kampala and its Suburbs.* Number 9 in East African Studies. East African Institute of Social Research, Kampala, Uganda.

Steinitz, C., Faris, R., Flaxman, M., Vargas-Moreno, J. C., Canfield, T., Arizpe, O., Angeles, M., Cariño, M., Santiago, F., Maddock, T., Lambert, C. D., Baird, K., and Godínez, L. (2005). A sustainable path? Deciding the future of La Paz. *Environment: Science and Policy for Sustainable Development*, 47(6):24–38.

Sui, D. Z. and Maggio, R. C. (1999). Integrating GIS with hydrological modeling: practices, problems, and prospects. *Computers, Environment and Urban Systems*, 23(1):33–51.

SURBANA International Consultants PTE Ltd. (2013). *Kigali Master Plan.* City of Kigali, Kigali, Rwanda.

Tatano, H., Yamaguchi, K., and Okada, N. (2004). Risk Perception, Location Choice and Land-use Patterns under Disaster Risk: Long-term Consequences of Information Provision in a Spatial Economy. In Okuyama, Y. and Chang, S. E., editors, *Modeling Spatial and Economic Impacts of Disasters*, Advances in Spatial Science, pages 163–177. Springer, Berlin, Heidelberg.

Tobler, W. R. (1979). Cellular Geography. In Gale, S. and Olsson, G., editors, *Philosophy in Geography*, volume 20 of *Theory and Decision Library*, pages 379–386. Springer, Dordrecht.

Turkington, T. A. R. (2016). *Changing flood and landslide hazard–A meteorological perspective*. PhD thesis, University of Twente, Enschede, The Netherlands.

Umer, Y. M., Jetten, V., and Ettema, J. (2019). Sensitivity of flood dynamics to different soil information sources in urbanized areas. Manuscript submitted for publication.

United Nations (2018). *World Urbanization Prospects: The 2018 Revision, Online Edition*. United Nations, Department of Economic and Social Affairs, Population Division, New York, NY.

van Leeuwen, P. J. (2009). Particle Filtering in Geophysical Systems. *Monthly Weather Review*, 137(12):4089–4114.

van Schrojenstein Lantman, J., Verburg, P. H., Bregt, A., and Geertman, S. (2011). Core Principles and Concepts in Land-Use Modelling: A Literature Review. In Koomen, E. and Borsboom-van Beurden, J., editors, *Land-Use Modelling in Planning Practice*, The GeoJournal Library Volume 101, pages 35–57. Springer, Dordrecht, The Netherlands.

van Vliet, J., Bregt, A. K., Brown, D. G., van Delden, H., Heckbert, S., and Verburg, P. H. (2016). A review of current calibration and validation practices in land-change modeling. *Environmental Modelling & Software*, 82:174–182.

van Vliet, J., Hagen-Zanker, A., Hurkens, J., and van Delden, H. (2013). A fuzzy set approach to assess the predictive accuracy of land use simulations. *Ecological Modelling*, 261:32–42.

van Vliet, J., Hurkens, J., White, R., and van Delden, H. (2012). An activity-based cellular automaton model to simulate land-use dynamics. *Environment and Planning B*, 39(2):198–212.

van Westen, C., Alkema, D., Feringa, W., Jetten, V., Hack, R., Brussel, M., Sliuzas, R., Charles-Griffiths, C., Hosein, T., Hazarika, M., Da Silva, R., Trigg, M., Sampson, C., Smith, A., Neal, J., and Hyde-Hecker, J. (2015). Caribbean Handbook on Risk Management. `http://www.charim.net/`. Online; accessed 9 Dec. 2015.

Veerbeek, W., Pathirana, A., Ashley, R., and Zevenbergen, C. (2015). Enhancing the calibration of an urban growth model using a memetic algorithm. *Computers, Environment and Urban Systems*, 50:53–65.

Verburg, P. H., Schot, P. P., Dijst, M. J., and Veldkamp, A. (2004). Land use change modelling: current practice and research priorities. *GeoJournal*, 61(4):309–324.

Vermeiren, K., Van Rompaey, A., Loopmans, M., Serwajja, E., and Mukwaya, P. (2012). Urban growth of Kampala, Uganda: Pattern analysis and scenario development. *Landscape and Urban Planning*, 106(2):199–206.

Verstegen, J. A., Karssenberg, D., Van Der Hilst, F., and Faaij, A. P. (2014). Identifying a land use change cellular automaton by Bayesian data assimilation. *Environmental Modelling & Software*, 53:121–136.

Verstegen, J. A., Karssenberg, D., van der Hilst, F., and Faaij, A. P. (2016). Detecting systemic change in a land use system by Bayesian data assimilation. *Environmental Modelling & Software*, 75:424–438.

Watson, V. (2014). African urban fantasies: dreams or nightmares? *Environment and Urbanization*, 26(1):215–231.

White, R. (1998). Cities and cellular automata. *Discrete dynamics in Nature and Society*, 2(2):111–125.

World Bank (2013). Project Information Document (PID) Concept Phase. Kampala Institutional & Infrastructure Development Project 2 (P133590). Technical Report PIDC1010., the Bank, Washington DC.

Xu, L., Gober, P., Wheater, H. S., and Kajikawa, Y. (2018). Reframing socio-hydrological research to include a social science perspective. *Journal of Hydrology*, 563:76–83.

Yeh, A. G.-O. and Li, X. (2001). A constrained CA model for the simulation and planning of sustainable urban form by using GIS. *Environment and Planning B*, 28(5):733–753.

Zischg, A. P., Hofer, P., Mosimann, M., Röthlisberger, V., Ramirez, J. A., Keiler, M., and Weingartner, R. (2018). Flood risk (d)evolution: Disentangling key drivers of flood risk change with a retro-model experiment. *Science of the Total Environment*, 639:195–207.