

MODELING URBAN GROWTH AND FLOODING INTERACTIONS WITH CELLULAR AUTOMATA IN KAMPALA, UGANDA

EDUARDO PÉREZ MOLINA
March, 2014

SUPERVISORS:

Dr. R. Sliuzas
Dr. J. Flacke



MODELING URBAN GROWTH AND FLOODING INTERACTIONS WITH CELLULAR AUTOMATA IN KAMPALA, UGANDA

EDUARDO PÉREZ MOLINA
Enschede, The Netherlands, March, 2014

Thesis submitted to the Faculty of Geo-information Science and Earth
Observation of the University of Twente in partial fulfilment of the requirements
for the degree of Master of Science in Geo-information Science and Earth
Observation.
Specialization: Urban Planning and Management

SUPERVISORS:

Dr. R. Sliuzas
Dr. J. Flacke

THESIS ASSESSMENT BOARD:

Prof. dr. ir. M.F.A.M. van Maarseveen (chair)
Dr. S. Geertman (External Examiner, Universiteit Utrecht)

Disclaimer

This document describes work undertaken as part of a programme of study at the Faculty of Geo-information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

“The ravages committed by man subvert the relations and destroy the balance which nature had established between her organized and her inorganic creations; and she avenges herself upon the intruder, by letting loose upon her defaced provinces destructive energies hitherto kept in check by organic forces destined to be his best auxiliaries, but which he has unwisely dispersed and driven from the field of action. When the forest is gone, the great reservoir of moisture stored up in its vegetable mould is evaporated, and returns only in deluges of rain to wash away the parched dust into which that mould has been converted.”

George Perkins Marsh. 1867. *Man and Nature*, pp. 43–44

ABSTRACT

A cellular automata model to simulate plausible land cover patterns for the city of Kampala, was developed, calibrated, validated and used to explore the impacts of policy actions and population growth. The model incorporates the changing impact of neighboring land cover, land cover at the base year (inertia), accessibility (estimated travel time to the CBD) and wetland areas, on the dynamics of land cover change. Flooding and hydrological outcomes are modeled by coupling the urban growth model to a LISEM flood model of the study area.

Existing land cover data models of Kampala, Uganda, for the years 2004 and 2010 were assessed and improved. Ancillary information was also extracted from compiled data sets in previous modeling efforts at ITC. This GIS was then used to develop the model and simulations.

Model calibration involved the calculation of auxiliary models: a set of models based on a single factor, to assess the effect of each potential determinant; the selection of a subset of potential factors; the comparison between different weighing schemes for this subset of factors, and the introduction of additional constraints (specifically, on institutional land uses). Each auxiliary model estimated en route to the final calibration was assessed, by visual comparison of the predicted patterns as well as a set of statistical measures of correspondence between both maps. The Upper Lubigi subcatchment, 2004-2010 period, was used as the calibration area. The final model (set of factors, associated weights and additional restrictions) was validated by applying it to the Nalukolongo catchment, 2004-2010. The model was found to be generally successful in replicating land cover patterns. The final predictions were more clustered than the land cover map, but ameliorated this problem by a better introduction of randomness, relative to previous simulation efforts.

Nine scenarios were simulated to explore the response of the land system to external disturbances as well as trend conditions. In addition to trend conditions (a slightly modified version of the calibrated model), four scenarios were assessed: the evacuation of areas flooded in 2010, a total ban on future development of wetlands, a cap on the percentage of development in each cell, and a lifting of trend constraints on wetland development. Three population growth rates were explored (trend, high growth rate as trend + 33% of trend and low growth rate, trend - 50% of trend). For each scenario, future land cover was predicted for 2020 (using 2010 with improved drainage system as a base line); the openLISEM flood model was used to estimate flooding and other hydrological characteristics. The total amount of development was found to be the key factor; the distribution of development has a minor impact, relative to runoff generation. Flood evacuation is judged as an inefficient strategy, except under low growth conditions.

Keywords

Urban growth, Cellular Automata, Flooding, Kampala, Uganda

ACKNOWLEDGEMENTS

First and foremost, I gratefully recognize the contributions of my supervisors, Dr. Richard Sliuzas and Dr. Johannes Flacke, who have been a constant source of support and knowledge on urban growth modeling in general and of Kampala. Without their advice during key stages of model development and their suggestions on scenario specification, this thesis would not have been completed.

Prof. dr. Victor Jetten kindly provided the calibrated flood model of Upper Lubigi, used in the calibration and simulation processes. His generous attitude as well as his willingness to discuss Kampala and its flooding problems, greatly improved the results obtained in this thesis.

The chairman of my proposal and mid-term committees, Prof. dr. Anne Van der Veen, contributed valued comments on scope, unclear aspects, as well as focusing my concerns on the most relevant issues.

During the project formulation stage of this project, I was fortunate to interact with a group of colleagues also wishing to work on Kampala. Valuable discussions with Garikai Membele, Epeli Nadraiqere, Esther Githinji, Gewa Li and lead by Dr. Walter de Vries, contributed to clarify my thinking and substantially improved the proposal document.

Ms. Siddhi Munde, dear friend and colleague from the GFM side of the divide, lent much appreciated advice on the remote sensing stage of this project. For this and for sharing her notes and material, I thankfully acknowledge her aid.

TABLE OF CONTENTS

Abstract	ii
Acknowledgements	iii
1 Background and Research Design	1
1.1 Motivation and Problem Statement	1
1.2 Conceptual Framework	1
1.3 Research Objectives	2
1.4 Methodological outline	4
1.5 Previous Work on land use and land cover modeling in Kampala, Uganda	4
1.6 Outline of the Report	5
2 Study Area Definition and Characterization	7
2.1 Demarcation of the Study Area	7
2.2 Land Cover Patterns in Lubigi and Nakulolongo	8
3 Cellular Automata Models in Urban Growth Modeling	11
3.1 Definition and Origins of Cellular Automata	11
3.2 Current Research Issues in Cellular Automata Modeling	12
3.3 Automaton State in Cellular Automata models	13
3.4 Meaningful Transition Rules in CA Models of Land Systems	14
3.5 Coupling of Flooding and Urban Growth Models	15
4 Cellular Automata Model Design and Estimation Strategy	17
4.1 CA Model Definition	17
4.2 Implementation Strategy	19
5 Cellular Automata Model Calibration and Validation	23
5.1 Procedure for Cellular Automata Model Implementation	23
5.1.1 Evaluation of predicted land cover patterns	25
5.2 Exploration of the Effects of Single Factors	26
5.3 Selection of Factors Combination and Weights	29

5.4	Use of Land Use Data to Refine Calibration Results	32
5.5	Cellular Automata Model Validation	34
5.6	Discussion of Calibration Process: Methodological Results	36
6	Scenario Development: Simulating Population Growth and Land Policy Interventions	39
6.1	Scenario Specification	40
6.1.1	Demographic Scenarios	40
6.1.2	Land Use Policy Interventions and Scenario Specification	41
6.2	Flooding Impacts of Simulated Scenarios	43
6.3	Discussion of Scenario Simulations	46
7	Concluding Remarks	49
7.1	Land cover modeling and simulation of scenarios	49
7.2	Methodological issues	50
7.3	Opportunities for further research	52
	References	53
A	Land Cover Data Model Development	59
A.1	Data inputs	59
A.2	Quality assessment of existing data models	59
A.2.1	Extent	59
A.2.2	Accuracy	60
A.3	Land cover improvement requirements	62
A.4	Land cover map derivation	63
A.5	Comparative accuracy assessment of land cover data models	64
A.6	Extension of 2010 land cover map	66
B	Generation of accessibility spatial index	69
C	Analysis of Land Cover Relationships of Built Up and Bare Soil Percentages, Neighborhood Effect and Other Inputs for Modeling	71
C.1	Simulation of Built Up Land Cover Demand Patterns	71
C.2	Neighborhood Effect: Definitions of Quantification and Interaction	73
C.3	Bare Soil and Built Up Land Cover	76

D	ModelBuilder Models: Algorithms for Land Cover Simulation	79
E	Statistical Assessment of Final Land Cover Model Results	81
F	Performance of CA Model	83
G	Simulated land cover for 2020 scenarios	85

LIST OF FIGURES

1.1	Conceptualization of structural relationships leading to land cover change processes in Kampala	3
2.1	Study area: calibration and validation	7
2.2	Land cover patterns of the Lubigi catchment, 2004 and 2010	9
2.3	Differences in the type of urban land uses and built up land cover between Upper Lubigi and Nalukolongo, 2010: typical examples	10
4.1	Schematization of the CA model: elements and calibration process for the full model	18
4.2	Rule for estimating bare soil percentage based on built land cover percentage in CA models	22
5.1	Land cover pattern predictions of the Upper Lubigi subcatchment, 2010. Single factor models	28
5.2	Land cover pattern predictions of the Upper Lubigi subcatchment, 2010. Multiple factor with equal weights models	31
5.3	Land cover pattern predictions of the Upper Lubigi subcatchment, 2010. Multiple factor with varying weights models	32
5.4	Land cover pattern predictions of the Upper Lubigi subcatchment, 2010. Final model and land use restrictions	33
5.5	Land cover pattern predictions of the Nalukolongo subcatchment, 2010. Final model and land use restrictions	35
6.1	Scenario Development Logic	40
6.2	Population estimates and projections for the city of Kampala	41
6.3	Spatial extent of spatial scenarios	44
6.4	Ratio of scenario value to 2010 conditions. Hydrological outcomes vs. average percentage of built up land cover	47
6.5	Ratio of scenario value to 2010 conditions. Flooded area vs. hydrological outcomes	48
A.1	Extent of input data models for land cover development	60
A.2	Land cover map derivation procedure for the study area, 2004	64
A.3	Land cover map of the study area, 2004	65
A.4	Area for which 2010 land cover map was extended	66

B.1	Flow chart for the derivation of travel time to CBD and subcenters	69
C.1	Histograms. Percentage of built up and its logistic transformation for cells in which built up percentages increased, Lubigi, 2004-2010	73
C.2	Change in the built up Percentage 2004-2010 vs. Neighborhood built up Percentage 2004. Scatterplot and smoothed trend (LOESS with bandwidth = 1/2) . . .	74
C.3	Change in the Bare Soil Percentage 2004-2010 vs. Neighborhood built up Percentage 2004. Scatterplot and smoothed trend (LOESS with bandwidth = 1/2) . . .	77
D.1	ModelBuilder Model: Algorithm to cycle through simulation periods and update neighborhood factor	79
D.2	ModelBuilder Model: Allocation algorithm to select through cells and identify which should change to fulfill exogenous demand	79
F.1	Model Performance: Number of Iterations in Allocation Routine vs. Percentage of Exogenous Demand Assigned for Final Calibrated Model	83
G.1	Predicted land cover maps for specified scenarios, 2020	85
G.2	Predicted land cover maps for specified scenarios, 2020	86

LIST OF TABLES

2.1	Land cover change in the study area, 2004-2010	10
5.1	Predicted built up land cover percentage, 2010, using single factor models. Comparison of predictions and land cover map	27
5.2	Predicted built up land cover percentage, 2010, using multiple factor models with equal weights. Comparison of predictions and land cover map	30
5.3	Predicted built up land cover percentage, 2010, of Nalukolongo using the final calibrated model. Comparison of predictions and land cover map	35
5.4	Pearson Correlation between factors and resulting 2010 built up land cover percentages	37
6.1	Demographic Inputs for Scenario Construction	42
6.2	Scenario specification: land demand and land use policy combinations	43
6.3	Hydrological Outcomes for Simulated Trend Scenarios, 2020	45
A.1	Confusion matrix. Accuracy assessment of land cover change in original building footprints, 2004-2010	62
A.2	Confusion matrix. Accuracy assessment of land cover change in building footprints as improved by Fura (2013), 2004-2010	62
A.3	Accuracy assessment of land cover data models. Overall accuracy and Kappa statistic	65
C.1	Descriptive statistics of built up percentage growth	72
C.2	Comparison of alternative explanatory demand models of logistic transformation of built up land cover growth	72
C.3	Comparison of predictive model of change in % of built up land cover, 2004-2010, as a function of average percentage of built up land cover in 2004 in moving window. Adjusted Coefficient of Determination	75
C.4	Comparison of predictive model of change in % of built up land cover, 2004-2010, as a function of average percentage of built up land cover in 2004 in moving window. Root Mean Square Error	76
C.5	Comparison of predictive model of change in % of built up land cover, 2004-2010, as a function of average percentage of built up land cover in 2004 in moving window. Akaike Information Criterion	76

E.1	Predicted built up land cover percentage, 2010, using multiple factor models with varying weights. Comparison of predictions and land cover map of calibrated model	81
-----	---	----

Chapter 1

Background and Research Design

1.1 MOTIVATION AND PROBLEM STATEMENT

The city of Kampala, as many other urban areas in the developing world, has recently evolved as a case of rapid growth resulting in an accelerated expansion of its urban footprint (Vermeiren et al., 2012). Because of a weak institutional setting but also due to a complex physical context, this expansion has generated a number of negative externalities - among them, increased urbanization has lead to greater runoff, a consequence of more impervious areas. This, in turn, has contributed to aggravating local flooding problems (Lwasa, 2010).

In the words of Cronshey et al. (1986, p. 1-1): “Urbanization changes a watershed’s response to precipitation. The most common effects are reduced infiltration and decreased travel time, which significantly increase peak discharges and runoff.” But the pattern of urban development itself has also been found to introduce substantial differentials (Mejía & Moglen, 2009; Yang et al., 2011).

The role of urbanization in increasing flooding justifies the interest on simulating urban development patterns, to assess their potential impact on the expected flooding and on the policy or structural solutions proposed to mitigate it. Previous studies of urban growth in Kampala (Vermeiren et al., 2012; Abebe, 2013; Fura, 2013) were based on the statistical examination of binary urban land cover change maps. They provide valuable information on the determinants of past land cover change as well as a basis for the generation of spatially explicit scenarios. However, the range of these scenarios is limited, as the models are highly influenced by past trends - particularly in the spatial conditions that determine them.

In view of this appraisal, the proposed research problem can be formulated thus: to create a simulation tool of land cover patterns that considers flooding within the modeling framework. This model must incorporate the specific characteristics of Kampala, Uganda, while retaining enough generality to be useful in other spatial contexts. Emphasis must be placed on the potential to create a diversity of meaningful scenarios which respond both to land use policy and other factors influencing the growth and distribution of population and other urban activities. Land cover simulation must also include a strong component of randomness to reflect weak enforcement of regulations and higher willingness to accept unsuitable conditions of land for development, notably of the urban poor.

1.2 CONCEPTUAL FRAMEWORK

Flooding in urban areas is becoming an increasingly important issue. Flood impacts are driven by a number of human factors, among them, land use and land cover partially determining flood frequency size and frequency (Jha et al., 2011). Physically, climate change and upstream environ-

mental degradation, which reduce the capacity of the landscape to absorb water or mitigate peak flows, are each generally contributing to increased flood hazards. But in the city itself, other factors may also force events, leading to greater impacts: urban development in flood prone areas reduces infiltration capacity in critical locations; decreased permeability of open spaces (i.e., city density increases - which is noted as positive for transportation but negative from a hydraulic view point); infrastructure problems develop over time, including poor maintenance, nonexistence or overload of drainage systems, and impacts of urban microclimates (heat islands). Other dimensions of risk, exposure and vulnerability, are also argued to be on the rise worldwide (Jha et al., 2011).

The practice of land use modeling includes a diversity of approaches, based on a limited amount of theories and methods (Koomen & Stillwell, 2007). Among them, the approach of Steinitz et al. (2003) (a more recent example of which can be found in Vargas-Moreno & Flaxman, 2012) is extraordinarily versatile, as it relies on organizing relevant information from disparate sources. In it, each location (cell in a regular grid) is assigned an attractiveness score, summarizing its desirability for a given land use (or its associated land cover). Certain cells are retired from the simulation (the so-called constraints) because they are deemed to be totally unsuitable for a given land use.

In the specific context of Kampala and of this research project, it is necessary not only to derive this overall suitability but also to quantify the contribution of diverse factors - especially the individual role played by flooding, because this factor is not exogenous to the system. Figure 1.1 summarizes the relationships considered: areas affected by flooding and areas occupied by wetlands are judged to negatively affect the development potential whereas areas with high accessibility (especially to the CBD) are taken to be desirable for urban development. For each of these factors, scores reflecting the degree of desirability (or lack thereof) must be assigned and combined using weighed summation into a suitability score, to determine which locations within the simulation change into built up land cover.

The changes in the land cover patterns would, through the runoff volume, affect the overall level of flooding. As more areas are urbanized, the degree of imperviousness increases (as does the runoff, the fraction of rainfall flowing over the landscape), resulting in more flooding. This added flooding may possibly change the suitability patterns: all other things being equal, the degree of suitability for urban development decreases and the location of the most suitable cells might also change. However, it is an open question whether this process considerably affects the distribution of development. For example, in terms of temporal scale, both process differ - although, of course, some catastrophic events may significantly impact urban patterns (as is clearly the recent case of New Orleans, among others).

1.3 RESEARCH OBJECTIVES

This project focuses on exploring the relationship between land cover patterns, particularly development density, and flooding. Its general objective has been defined as: *to develop a model of built up density variations in space that would be coupled with a flood model and used to generate scenarios on potential reactions of flooding to land cover patterns.*

A cellular automata (CA) model was selected to fulfill these requirements¹. Five specific objectives and a number of research questions, related to the building of this CA modeling tool in the specific context of Kampala, were postulated:

¹CA was thought appropriate because of local neighborhood effects have been detected to be very influential in explaining past urban development in Kampala. The empirical evidence from previous statistical studies is discussed in section 1.5.

1.4 METHODOLOGICAL OUTLINE

The research project was developed in three successive stages: a first consisted of an in-depth description of the study area (the catchments of Lubigi and Nalukolongo in Kampala). This description, in turn, includes the assessment of previously existing land cover maps, the development of new and improved land cover maps, the generation of spatial indices reflecting accessibility and restrictions due to wetlands, and flooding maps based on current land cover conditions. All of these intermediate products were then used as inputs of the modeling process, as well as in informing the design of the model.

A second stage included the design and calibration of the CA model, for the Lubigi catchment and the period 2004-2010. The parameters derived were then used to validate the model by predicting land cover for 2010 in Nalukolongo, from a 2004 land cover map. An overarching problem in this step was the evaluation of predicted results, which in turn was used to examine the contribution (in terms of a better prediction) of introducing new elements.

Finally, in a third stage, scenarios of future land demand (based on expected population growth for the entire region) and land policy were specified. The model was then used to estimate future land cover and flooding for each combination of population and policy that constituted a scenario.

1.5 PREVIOUS WORK ON LAND USE AND LAND COVER MODELING IN KAMPALA, UGANDA

The general determinants of urban development are known from the various analytical traditions, particularly on residential location (the model of Alonso, 1964, and further extensions, see synthesis in Brueckner, 1987) and environmental planning (suitability analysis, as proposed in McHarg, 1969 and operationalized, among others, by Steinitz and colleagues, e.g. Steinitz et al., 2003). Within the CA modeling field, the SLEUTH model (Clarke et al., 1997) makes use of a relatively standard set of spatially explicit criteria, which have been found to be applicable in a wide variety of urban settings (Clarke et al., 2007), these factors being slope, land use, urban, exclusion, transportation and hill shading.

Previous analysis of the city of Kampala have already quantified the impact of the most important determinants of its urban morphology, at mid and detailed scales. These approaches are all statistical (making use of Logit econometrics, without spatial autocorrelation but including at least a proxy of neighborhood effects). Three of the previous studies (Abebe, 2013; Fura, 2013; Mohnda, 2013) were developed in the context of an ITC project evaluating the impact of flooding on the built up area (Mohnda, 2013, is a flood model that incorporates land cover data but no modeling of LUC; see Sliuzas et al., 2013, for a general description of the project aims and integrated results). A fourth (Vermeiren et al., 2012) preceded them. All show broadly consistent results.

A notable common feature of these spatial-statistical models is the inclusion among the determinants of variables representing neighborhood effects. Measures to mitigate problems of spatial autocorrelation through sampling are also common (Vermeiren et al., 2010, creates a random sample; Fura, 2013, follows the precautions taken, among others, in Cheng & Masser, 2003) albeit none of these papers implement existing econometric tools such as Probit Bayesian models with spatial autocorrelation (LeSage & Pace, 2009) or the original frequentist version proposed by McMillen (1992).

Fura (2013) included among the determinants of his model the variable *proportion of urban land in the surrounding area*, defined for a cell as the percentage of urban cells within a radius of 7 cells

centered on it and computed for the base year 2004 (the Fura model was calibrated with data from the period 2004-2010). This variable had by far the largest odds ratio, exceeding by one order of magnitude the second largest odds ratio of the other variables. The models were based on land cover maps of building footprints derived from high resolution imagery.

Abebe (2013) estimated three models for successive pairs of years, again using the proportion of urban land within a 7 cell radius as a neighborhood effect. The odds ratio of the determinants for all three models was the greatest, even by a larger margin than in Fura (2013). These land cover maps were derived from mid resolution remotely sensed imagery (TM and ETM+ sensors carried by several of the Landsat satellites).

In the work of Vermeiren et al. (2012), a neighborhood effect was introduced by estimating a *built up potential* measure, defined for a cell as the sum of the inverse squared distance from all other built up cells to the cell in question. As in Abebe (2013), land cover data were also interpreted from imagery produced by the Landsat project.

The strength of these predictors suggests that the neighborhood of the locations has an important influence on whether a location is converted to built up land cover or not. This characteristics implies that CA - which relies on the state of neighboring cells to determine changes in the state of any given location - is a promising technique to improve the simulation of land cover patterns in Kampala.

1.6 OUTLINE OF THE REPORT

This thesis is divided into five sections, additional to these introductory remarks and the concluding chapter. A second chapter describes the study area, in terms of its extent and land cover change. The third chapter summarizes the theory and background of CA models applied to urban studies, with particular emphasis on the state of the automata, as well as other issues theoretically relevant to the modeling effort undertaken, specifically the joint modeling of flooding and urban growth.

The fourth chapter includes the design of the model as well as the implementation strategy. Chapter five summarizes the calibration and validation processes - including a discussion on the substantive results that were derived from them. Chapter six reports on the use of the calibrated model to simulate future land cover patterns; it includes the specification of the simulated scenarios and the overall logic of their construction, the results of the simulation and a discussion of them. Finally, a concluding section synthesizes the results achieved.

A series of appendices provide an expanded view over specific technical issues. Appendix A documents the assessment of available land cover data and raw imagery, and the generated land cover maps for the study region. Appendix B includes the flow charts of the ArcGIS 10.1 ModelBuilder Models used in generating accessibility and physical suitability (the effects of wetlands). Appendix C documents the estimation of the neighborhood effect, through a moving window of varying sizes, and its empirical relationship with the percentages of built up and bare soil land cover in each cell. Appendix D includes a schematic view of the algorithms used to model the process of demand allocation. Other appendices include the statistical assessment of the fully calibrated land cover model predictions (appendix E), an analysis of the relationship between computation time and number of iterations in the allocation model (the allocation component's precision, appendix F) and the land cover maps resulting from the simulations specified in chapter 6 (appendix G).

Chapter 2

Study Area Definition and Characterization

2.1 DEMARCATION OF THE STUDY AREA

The study area in which this case study was developed includes two subcatchments located within the city of Kampala, Uganda (see figure 2.1). They have a combined area of 44.2 km^2 (63.4% is occupied by Upper Lubigi and 36.6% by Nalukolongo). Both correspond to the upper reaches of the hydrological system and drain inland, as opposed to directly into Lake Victoria. They also occupy peripheral but important subcenters of the city.

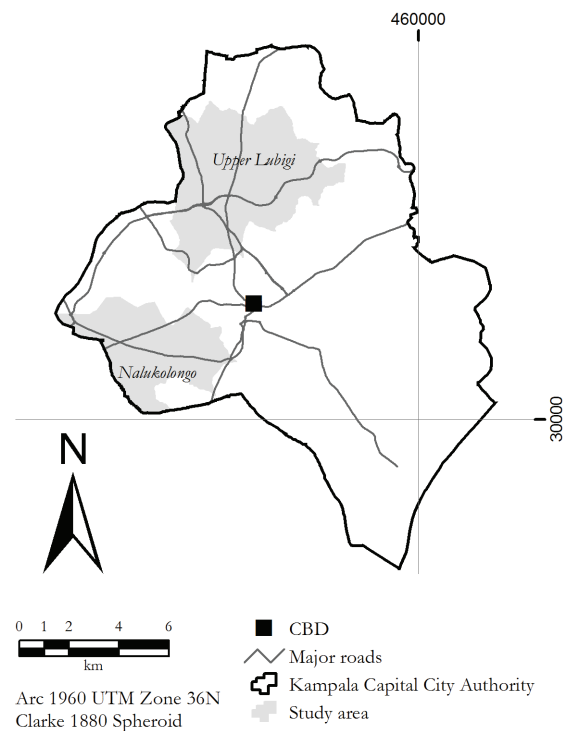


Figure 2.1: Study area: calibration and validation

Previous research efforts on modeling urban growth and its impact on flooding patterns, performed at ITC as part of an integrated flood management project (Sliuzas, Lwasa, et al., 2013), have concentrated on the Upper Lubigi subcatchment. Within this overall project, information to derive land cover data models in two periods is available. The lack of a third period has precluded model validation efforts; thus, the second subcatchment (Nalukolongo) was added.

The strategy defined was to formulate and calibrate the model using only data from the Upper Lu-

bigi subcatchment. Afterwards, data from Nalukolongo was used to validate the derived model. To ensure consistency, land cover data models were created by interpretation of the available imagery (see appendix A).

The original research design included an expanded calibration area, the entire area of the Lubigi subcatchment that is located within the KCCA. Indeed, land cover data models developed for 2004 and 2010 (reported in Appendix A, see figure A.3) include all of the Lubigi and Nalukolongo subcatchments and the analysis of the optimal neighborhood size also included all of Lubigi. However, since flood depth was tested as a potential explanatory factor and no soil permeability data was available outside of the Upper Lubigi area, the final calibration was restricted to Upper Lubigi.

2.2 LAND COVER PATTERNS IN LUBIGI AND NAKULOLONGO

Land cover patterns for the study area for the years 2004 and 2010 are reported in figure 2.2. The process of urban growth can be characterized by the observed land cover changes: both in Upper Lubigi and Nalukolongo, the 2010 pattern is darker than the 2004, which is to say, new buildings have been constructed. More development occurred in Upper Lubigi, clearly; but it is also notable that, in both cases, much of the land cover change into built up took place in areas with previously little building (this is even more evident when comparing the patterns of vegetation land cover, and how the central parts of both subcatchments lost vegetation cover).

Bare soil also increased generally, and resulted in a more dispersed pattern in both subcatchments. It seems more clustered in Nalukolongo than in Upper Lubigi, but in both cases, it is similar to the resulting pattern of built up land cover (suggesting a link between them). Perhaps the differences between Nalukolongo and Upper Lubigi - more grouped pattern of bare soil and less growth of built up in Nalukolongo - may suggest a less dense pattern of urbanization in Nalukolongo, relative to Upper Lubigi. But other alternative explanations may also exist.

It is interesting to note that the rates of growth in built up land cover are similar for both subcatchments but, in Nalukolongo, the bare soil category growth rate doubles that of Upper Lubigi. Perhaps the larger regional accessibility of Upper Lubigi - as witnessed by the larger density of main roads in this subcatchment (see figure 2.1) - has been compensated by a greater increase in local accessibility, required to properly urbanize areas of Nalukolongo.

It is also important to note that the tarmac land cover - main roads - increased for Upper Lubigi and essentially remained constant in Nalukolongo (the small reduction in the tarmac land cover is probably due to uncertainty in the land cover data models). This not only reinforces the preceding argument; it also contributes to explain why Upper Lubigi has larger levels of urban growth (slightly higher rate of growth for built up land cover but with a larger area built up in 2004, and also higher density of development). Indeed, it is to be expected that the more accessible option will be more intensely used for urban land uses.

Finally, it is also important to note a difference in the *urban land uses* between both areas, which might introduce differences in the model's predictions. Both catchments are fundamentally occupied by detached single housing units, a large proportion of which is informal (this can be seen in the southern parts of both the Upper Lubigi and Nalukolongo zooms in figure 2.3, with small buildings located with little regularity of orientation). Upper Lubigi also includes: (a) some buildings, either residential or institutional (as can be seen in the western edge of the zoomed image, figure 2.3), (b) a very important part of the system of regional roads. Nalukolongo also has access by the regional road system but less so (roads and intersections seem to have lower capacity);

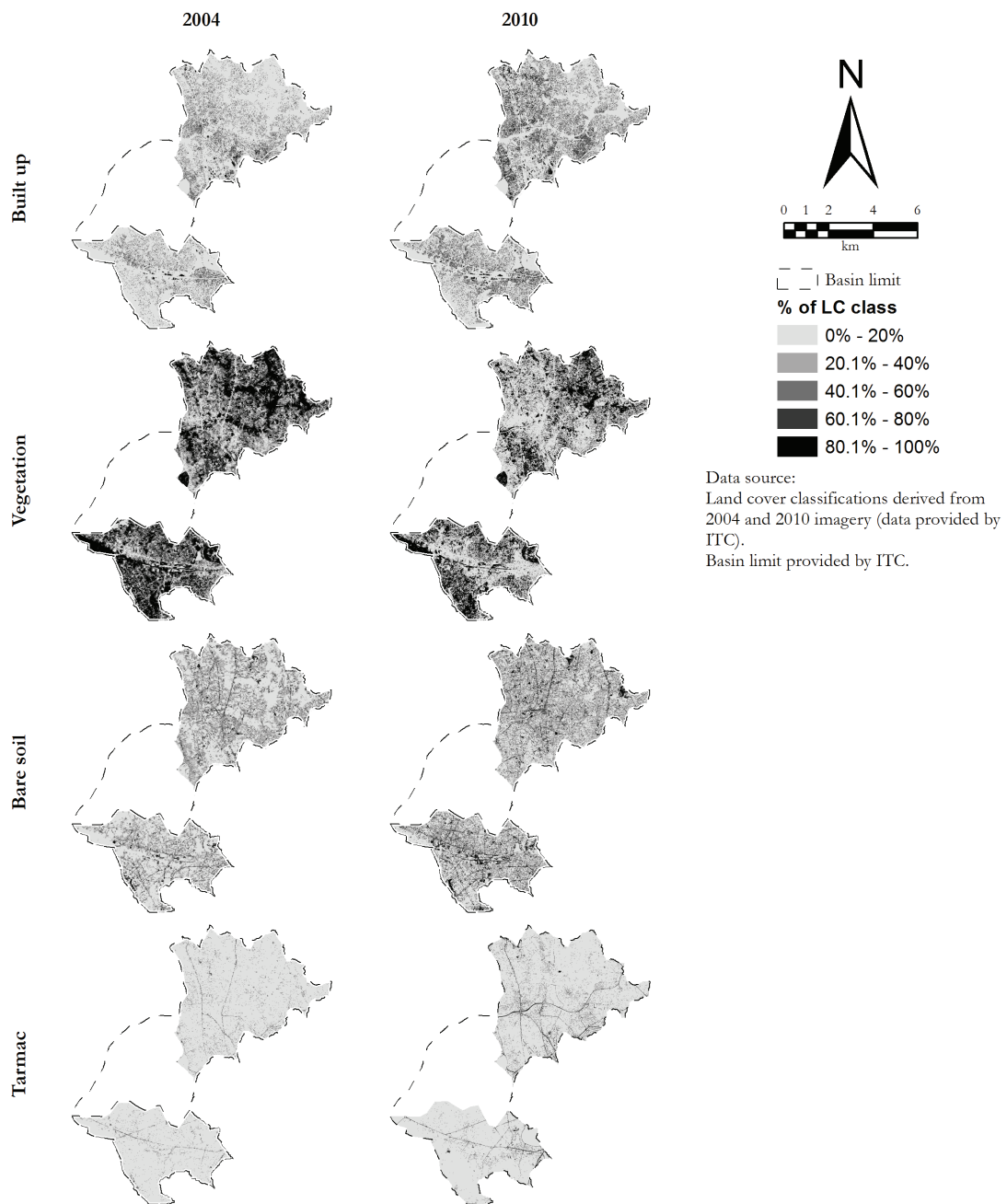


Figure 2.2: Land cover patterns of the Lubigi catchment, 2004 and 2010

further, the large buildings are evidently different: they are buildings for industrial use, whether logistic (warehouses) or productive (manufacture). This type of development is typically created in the periphery because it requires large areas - for freight vehicle operations, future growth, etc. It is also, and in consequence, less dense (meaning the small residential detached housing cannot occupy the small areas between these buildings or even adjacent to them).

Table 2.1: Land cover change in the study area, 2004-2010

	2004		2010		Inter-annual Growth Rate
Land cover	Area (ha)	%	Area (ha)	%	
Upper Lubigi					
Built up	140.9	15.4%	212.6	23.3%	7.1%
Vegetation	482.9	52.9%	343.2	37.6%	-5.5%
Bare soil	202.8	22.2%	249.2	27.3%	3.5%
Tarmac	85.9	9.4%	107.8	11.8%	3.8%
Water	0.1	0.0%	0.2	0.0%	7.1%
No data	0.1	0.0%	0.0	0.0%	
Total ^{1/}	912.8	100.0%	913.0	100.0%	
Nalukolongo					
Built up	70.3	13.1%	103.3	19.2%	6.6%
Vegetation	310.0	57.7%	238.5	44.4%	-4.3%
Bare soil	104.5	19.5%	152.3	28.3%	6.5%
Tarmac	42.0	7.8%	40.7	7.6%	-0.5%
Water	2.7	0.5%	2.7	0.5%	0.0%
No data	7.9	1.5%	0.0	0.0%	
Total ^{1/}	537.4	100.0%	537.4	100.0%	

1/ Totals do not match because of rounding of areas per category and border effects in the land cover data models.

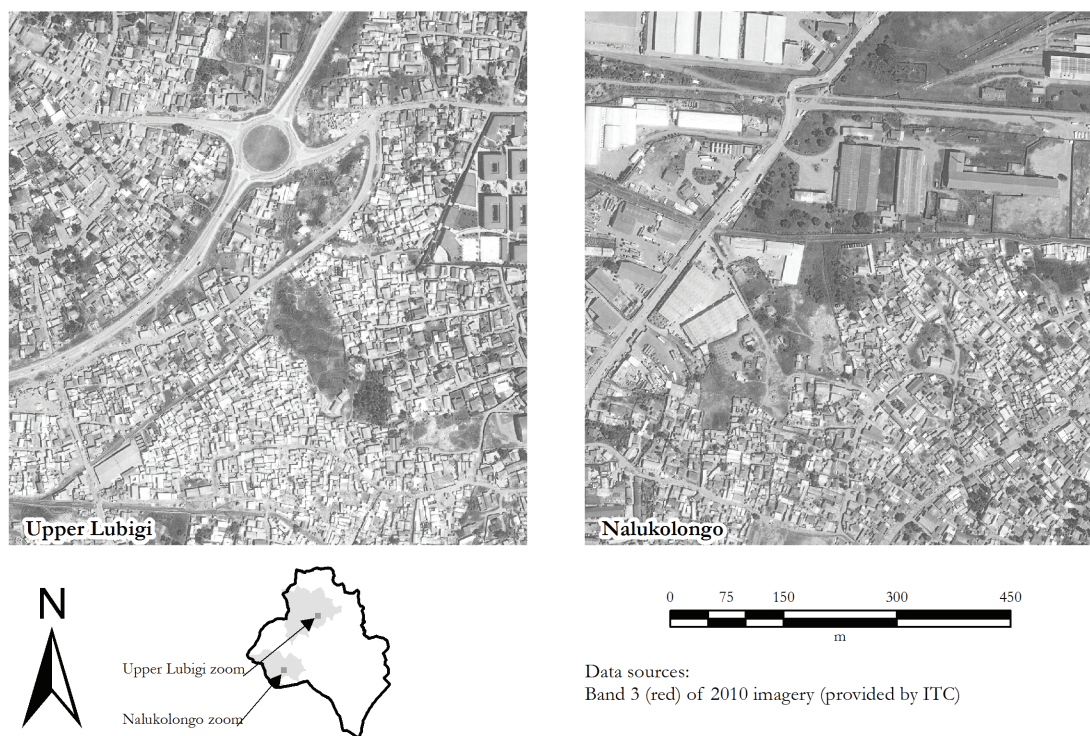


Figure 2.3: Differences in the type of urban land uses and built up land cover between Upper Lubigi and Nalukolongo, 2010: typical examples

Chapter 3

Cellular Automata Models in Urban Growth Modeling

3.1 DEFINITION AND ORIGINS OF CELLULAR AUTOMATA

CA can be defined as “a system of spatially located and interconnected automata” (Benenson & Torrens, 2004, p. 5). Each automaton is characterized by a state (according to Santé et al, 2010, this state can be a binary value, a qualitative value representing different land uses, a quantitative value symbolizing a feature of the land use or a vector of attributes). Further, “[t]he state of each cell depends on its previous state and on the state of its neighboring cells according to a set of transition rules” (Santé et al., 2010, p. 109)¹.

CA were originally conceived by John von Neumann in the early 1950s, based on discussions with Stanislaw Ulam and also likely influenced by the ideas of Allan Turing on what he called *automatic machines* (Benenson & Torrens, 2004, p. 95; Toffoli & Margolus, 1987, p. 9). Benenson and Torrens also cite early advances in Cybernetics by McCulloch and Pitts on logic and the transmission of information between neurons, and by Wiener and Rosenblueth on excitable media. The latter is particularly important for the field of spatial simulation because it applied the diffusion process to a space conceived as a collection of discrete entities².

Benenson and Torrens trace the introduction of CA into geographical modeling of urban patterns as arising from early developments conceptualizing land patterns as raster computer maps (in particular, three models: of Greensboro, NC, by Chapin and colleagues; of Boston, MA, by Steinitz and Rogers, and of Buffalo, NY, by Latrop and Hamburg, all in the United States). Some features exhibited in these models include the representation of urban data by means of multiple raster layers (as in Steinitz and Rogers), and the differentiation between the potential of a cell for a certain land use and its actual allocated land use (i.e., its state). These models all conceived space as an array of cells, each being characterized by a state, and they were all dynamic (cell states were updated at each time step). They do not, however, link the state of a cell with the state of its neighboring cells - the distinguishing feature of CA models.

This last gap would eventually be filled by Tobler (1979), with a very simple but conceptually powerful typology of land use models, which included: the independent model (land use in a given location is the result of randomness), the dependent model (land use in $t + 1$ depends on land use in t), the historical model (land use in t depends on land use, at that same location, during several preceding periods), the multivariate model (land use at the location depends on a set of characteristics of that location), and the geographical model (land use at the location depends on

¹Transition rules are a “set of conditions or functions that define the state of change in each cell in response to its current state and that of its neighbors.” Maithani (2010)

²“Excitable CA attempt to reduce an excitable medium to its simplest possible form. Each cell in an excitable CA is connected to some fixed number of neighbors. The simplest excitable rule has three states, 0, E , and R [...]. If a cell is in E then at the next time step it becomes R and after that it is returned to 0. If a cell is in state 0 and at least one neighbor is in state E , the cell is put in state E .” (Ermentrout & Edelstein-Keshet, 1993, p. 99)

the land use of surrounding locations). What is particularly elucidating of this theoretical work, is Tobler's contention that the five models are *simple* abstractions from nature, which could be combined to obtain more realism but that such process would be complicated (the second half of this paper develops the geographical model, which is evidently a pure CA model).

Eventually, advances in computational technology allowed for the implementation of the theory of CA modeling. A wealth of applications have been developed since the 'explosion' of this subject in the literature that occurred in the 1990s (Benenson & Torrens, 2004). However, applications seem to have emerged in a haphazard way, in no small part because of the ease and aptness of the method for simulation of spatially explicit patterns.

3.2 CURRENT RESEARCH ISSUES IN CELLULAR AUTOMATA MODELING

Modeling with CA greatly enhances the capabilities of geographic information systems (GIS). CA can integrate spatial and temporal dimensions of urban processes (Santé et al., 2010; Maithani, 2010), thus extending the domain of traditional GIS (Maithani, 2010). CA are regularly used, in the context of landscape modeling, to simulate trends (particularly, to replicate known past land cover scenarios in calibrating a model) but also alternative scenarios of future landscape, including the optimization of landscape patterns subject to a policy goal (Li, 2011). This expansion of GIS is achieved through simple, and consequently easily comprehended, means (Santé et al., 2010). Simplicity, however, can also pose grave limitations to the practical applications of CA (Santé et al., 2010; Timmermans, 2012). A criterion to resolve the degree of simplicity vis-à-vis its realism should be an important feature of any such modeling effort.

Santé et al. (2010) identify the absence of a standard for transition rules as a condition of particular importance. One of the promising areas of future developments designated by them is, precisely, the work on systematic calibration strategies, as current practice relies mostly on expertise and adjustment to specific cases (Li, 2011) or on blind and meaningless numerical optimization.

As indicated by Li (2011), solicitation of transition rules is challenging, since the modeled patterns exhibit inherent unpredictability. Automatic methods to define the transition rules have been used (Santé et al., 2010; Li, 2011): logistic regression, artificial intelligence (neural networks), and genetic algorithms, such as those applied in the SLEUTH (Clarke et al., 1997) and DINAMICA (B. Soares-Filho et al., 2013) models. In particular, potential exists for the use of logistic regression in the definition of empirically elicited transition rules (Fang et al., 2005). But any such strategy must not sacrifice a material interpretation of the relationships for links which are meaningless in themselves, even if they result in better predictions - as this would yield too much on the simplicity for a greater accuracy of prediction *that could well be the product of chance or bias*, rather than substantive improvements in the model.

That being said, it is quite evident that the original simplicity of CA models is too coarse to adequately represent land cover patterns in real applications (Santé et al., 2010; Timmermans, 2012). Relaxation of these early conditions, such as extending the neighborhood concept (Timmermans, 2012) or introducing ancillary spatial factors to account for variations in space (in the vein of the models originally proposed by White & Engelen, 1993), is already standard practice in CA modeling.

The evaluation of model results is also an area which requires careful specification. Li (2011, p. 394) formulates the issue thus: "Should these models be calibrated [and evaluated] by spatial validation (e.g. using overall accuracies and kappa coefficient) or by aggregate validation (e.g. using landscape

pattern metrics)?” Santé et al. (2010) are similarly conscious of the importance of this subject, when citing new validation strategies based on urban pattern recognition as a field for further inquiry. Timmermans (2012) criticized the emphasis on overall pattern assessment at the expense of the underlying process, which is rarely if at all appraised. But one should question, at least in the field of urban growth modeling, if it is reasonable to expect a successful prediction for any specific location. An intermediate solution was proposed by van Vliet et al. (2013), who defined a “fuzzy Kappa” that adds degrees of similarity to the computation of the traditional Kappa, better reflecting the overall pattern by means of a pixel-by-pixel evaluation.

3.3 AUTOMATON STATE IN CELLULAR AUTOMATA MODELS

Most applications of urban growth modeling have conceived space as regular tessellations with the state of each cell a discrete category. Typically, these have been land use or land cover classes. However, there is no inherent reason for this other than the fact that traditional data models of land use and land cover have been designed with this formulation.

Previous work exists on models with cells having a continuous variable describing their state. Jain (2009) contains an interesting conceptual discussion on the definition of the state for a CA model reflecting variations not only on development density - conceived in 3D (i.e., as volume of development) - but simultaneously on land use. The strategy discussed was to transform quantitative statements into categories. Cell state is characterized by the combination of three variables’ level: land use, which is *potential* (possible in the regulation) rather than dominant land use, job activity and density of buildings. Four intervals (and four land use classes) of each are used, resulting in 52 categories (because residential land uses have, by definition, no jobs and therefore a single category of job density). The remarkable thing in the model design is how this complex set of potential states transition, with only four rules for upgrading simultaneously development and jobs densities, by evaluating separately at each time step the relative potential for land use, jobs and development.

Li & Yeh (2000) and Yeh & Li (2001) proposed an alternative definition, which adopts a continuous variation as the state of a cell in CA models, termed by them grey cells. Their original aim included two goals, the linking of CA models with GIS (which at that point was incipient) and to expand the concept of the state of the cell. In their own words, “[t]he state of a cell in the ‘grey-cell’ method is expressed by a continuous value for development or conversion. The value indicates the cumulative degree of development for a candidate cell before it is completely selected for development or conversion.” (Yeh & Li, 2001, p. 736). The grey cell temporarily stores the cumulative change in cell suitability caused by changes in its neighborhood which, beyond a certain threshold, were then considered as developed (this suitability is assumed to increase with each time step). In their prior work, the grey cell represented “the degrees or percentages of urban land development during the iterations of modelling” (Li & Yeh, 2000) - a formulation clearly useful when addressing the impacts of land cover on flooding.

The use of a continuous variable to represent the state of a cell in a CA has few applications. Notable among them, Yeh & Li (2002) follow up on their earlier work and use the grey cell concept to model population densities with CA. They limit the amount of residents each cell can accommodate and they allocate total demand subject to total population growth (which is to say, they allocate residents). This formalization allows them to define a negative exponential decay function on the population density (a spatially varying global constraint).

The grey cell idea was generalized by van Vliet et al. (2012) to incorporate ‘activities’ in a CA

model: the state of the cells was characterized by them using two dimensions, a predominant land use category but also a vector storing diverse numeric characteristics of the cell. In this specific case, the vector stores quantitative data on two continuous dimensions, jobs and population (the case study is a stylized, two activity CA model developed to test the concept of activity and land use as state descriptors of a cell). The strategy followed was to define in each iteration the potential for all activities, to update the amount of each activity and - based on the comparison of updated activity levels - to classify the cell according to its predominant land use (in this case study, residential or industrial). White et al. (2012) used the same concept of activity based CA model but extending the neighborhood to the entire study area, “they represent interaction effects at all spatial scales; effectively, they are distance decay functions” (White et al., 2012, p. 1251).

3.4 MEANINGFUL TRANSITION RULES IN CA MODELS OF LAND SYSTEMS

As discussed in section 3.1, Tobler’s (1979) vision of a generalized model to explain an automaton’s state includes up to five different relationships, only one of which is the influence of the immediate surroundings. It was, thus, an evident next step to expand pure CA models to consider other factors. In particular, the combination of an independent (random) model, a multivariate model (of the influence of the characteristics of a location on its potential for a specific land use, which is to say, its suitability) and the geographical model (the term used in Tobler’s 1979 to refer to the neighborhood effect) was proposed by White & Engelen (1993).

White and Engelen’s formulation was summarized in White (1998). Four elements were included in the CA model’s transition rules: “intrinsic suitabilities representing inhomogeneities in the geographic space being modelled (e.g. soil quality or legal restrictions)” and “a local accessibility effect, representing ease of access to the transport network”, both of which jointly determine the multivariate model; “the neighbourhood effect, representing the attractive or repulsive effects of the various land uses (cell states) in the neighbourhood [...] of the target cell” (in Tobler’s parlance, the geographic model), and “a stochastic perturbation capturing the effect of imperfect knowledge and varying needs and tastes among the implicit actors whose decisions are represented by cell state transitions” (White, 1998, p. 115). Additionally, this model includes a global constraint, meaning that the total number of cells that change into a new state (land use category) are exogenously determined. Thus, the *location* of the cells that change is defined within the CA model but the amount of change is externally determined. In this sense, the White-Engelen models are equivalent to the early raster map models discussed in 3.1, with the addition of a neighborhood effect among the determinants of the development potential.

This same issue was addressed, in approximately the same manner, in the work of Wu (1998) - he formulates the derivation of the development score as the application of the method of analytical hierarchical process of multicriteria evaluation. In related work by Wu & Webster (2000), the more heuristic-style transition rules were replaced by theoretical micro-economic models which allow the incorporation of spatial externalities and of densification. Results achieved a long-run equilibrium in the market but with a continuously shifting spatial configuration. This latter piece is important because it linked a strong theoretical model, which can be used to construct and predict the consequences of policy and other exogenous influences, to their spatial manifestations - this link achieved through the CA model.

The patterns which result from simulation through CA models are an emerging property (Batty & Torrens, 2005), which may be (or not) the result of ‘sensible’ transition rules. The crucial point of the White-Engelen approach, eventually developed as the METRONAMICA software (van

Delden et al., 2005), is the specification of each one of these interactions. Since there may be many such transition rules - especially when transition rules consider pairwise impacts between uses, e.g., the influence of manufacturing over residential and vice-versa -, the calibration procedure relies on including only materially significant interactions, i.e., only those corresponding to real-world dynamics.

The work of Wu & Webster (2000) attempted to go beyond this solution by allowing for the determination of the impact caused by an external disturbance. The the spatial pattern emerging from the micro-economic model would have been different, given the inclusion of a specific exogenous disturbance in the micro-economic part of the model. It did not succeed in the sense that the resulting spatial pattern being stable in the long run; therefore, it was not possible to associate a long run spatial pattern in equilibrium to the specific disturbance causing it. Nevertheless, this approach does point to a limitation in CA modeling practice to date. The White-Engelen approach is useful in the sense that transition rules consistent with reality would presumably lead to land patterns better reflecting this reality. But while it is possible to construct models that result in more realistic patterns and with transition rules consistent with real-world processes, it is very difficult to isolate the contribution of each factor within these transition rules.

3.5 COUPLING OF FLOODING AND URBAN GROWTH MODELS

Despite it being a relatively straightforward approach, coupling of urban growth and flood models has been relatively rare, perhaps because, until recently, hardware has been a limitation on the development of applications. In general, most case studies have analyzed the impact of land patterns on hydrological or hydraulic impacts. Fairly typical examples of recent work include Ciavola et al. (2012), Huong & Pathirana (2013) and Poelmans et al. (2011). Applications introducing flooding as a restriction on land cover change are less common. The work of Steinitz and colleagues follow a flexible strategy that allows the introduction of many factors relevant to specific cases; in their application to La Paz, Mexico (Steinitz et al., 2005), a buffer to frequently flooded but usually dry watercourses, was retired from development for the more environmentally restrictive scenarios.

Urban growth modeling has relied heavily on the extrapolation of land cover trends. Ciavola et al. (2012) modeled economic and demographic trends to estimate demand and allocated this demand with the SLEUTH model (Clarke et al., 1997). The predicted scenarios were then input into environmental models - the American Environmental Protection Agency's Hydrological Simulation Program - FORTTRAN (HSPF) - to estimate hydrological and nutrient outputs. Huong & Pathirana (2013) simulated future land cover trends with the Dinamica EGO model (B. S. Soares-Filho et al., 2002), jointly with an atmospheric forcing model; unlike the Ciavola et al. (2012), Huong & Pathirana (2013) did not estimate changes in runoff. Rather, land cover simulations were used to explore the sensitivity of atmospheric conditions to land systems; hydraulic flood routing assumed a single condition for urban land use (results were also combined with sea level rise and riverine flood increases due to climate change). Poelmans et al. (2011) do use a runoff model, specified and calibrated for their specific case study; their hydraulic model conceives floodplains as reservoirs and jointly calibrates the parameters determining the flood depth with runoff (flood depth is calculated with the InfoWorks RS software).

A characteristic of this coupling is the combination of information at multiple spatial and temporal scales. Ciavola et al. (2012) predict socioeconomic demand at county level, disaggregate information within counties with uniform pixels and aggregate the final outputs per watershed. Outputs of the work by Huong & Pathirana (2013) were spatially explicit flood depths, separated

according to their primary cause (river or rainfall). Poelmans et al. (2011) also produced results per subcatchment.

Temporal scales represent an additional important limitation, especially when climate change is incorporated into the modeling. For example, Huong & Pathirana (2013) simulated a 100 year period, which clearly exceeds the possibilities of urban growth modeling (particularly if these are based on historical trends). The use of scenarios in this field is required due to the very large uncertainties introduced by multiple scales.

Eigenbrod et al. (2011) is a different case study, a nation-wide simulation of Great Britain built with the objective of showing the tradeoff in ecosystem services of dense urban development (resulting in increased flooding risk) and suburban development (which generates large losses of carbon storage and agricultural production). The spatial resolution is very coarse (1 km cells) yet the results of the simulation of both urban growth and ecosystem services, including flooding, are plausible.

Regarding Kampala, the LISEM model - originally proposed by De Roo et al. (1996) - has been calibrated for the city. It was used by Mohnda (2013) to assess various runoff reduction strategies in the Lubigi catchment. More generally, the results of Logistic urban growth models developed by Fura (2013) were used as inputs for the openLISEM model to estimate a diversity of scenarios which included both future plausible land cover (as estimated by Fura) as well as interventions on the drainage system and alternative infiltration actions (Sliuzas, Flacke, & Jetten, 2013). openLISEM results of these assessments include the estimation of the volume of water infiltrated, intercepted and runoff, as well as the outflow estimate, and a hydrogram that can be used to analyze the peak discharge (Mohnda, 2013). The most recent versions of openLISEM also produce spatially explicit simulations (in time) of the changes in flood depth (Jetten, 2013; Sliuzas, Flacke, & Jetten, 2013).

Chapter 4

Cellular Automata Model Design and Estimation Strategy

4.1 CA MODEL DEFINITION

A CA model has been formalized to simulate land cover scenarios in the city of Kampala and examine their consequences on flood impacts.

Within the model, the study area is divided into a regular grid of square cells. Each cell is conceived as an automaton, characterized by a set of states ($\mathbf{G}$), a set of transition rules ($\mathbf{T}$) governing changes to this state and a set of states of neighboring cells ($\mathbf{R}$):

$$Au(\mathbf{G}, \mathbf{T}, \mathbf{R}) \quad (4.1)$$

with Au the automaton. The state $\mathbf{G}$ of the automaton Au is not discrete; it was defined as a set of two continuous variables representing the percentage of built up area, $G(1)$, and the percentage of bare soil, $G(2)$, present within that cell. Both are bounded values, $G(l) \in [0, n]$, $0 < n < 1$, $l \in \{1, 2\}$.

The transition rules $\mathbf{T}$ define the state of the automaton ($\mathbf{G}_{t+1}$) in time step $t + 1$, 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):

$$\mathbf{T} : (\mathbf{G}_t, I_t) \rightarrow \mathbf{G}_{t+1} \quad (4.2)$$

In a classic cellular automata model, the input I_t would be solely determined by the state of the cells in the neighborhood¹ of the automaton ($\mathbf{R}$).

The full model proposed is schematized in figure 4.1, where it can be seen that the input, I_t , is defined by two components: a composite of factors that affect the desirability of land for building (i.e., its development potential) and a neighborhood effect, in the sense of the classic CA model.

The composite was created by the weighed summation of the input layers of ancillary data, which represent what White (1998, p. 115) calls “intrinsic suitabilities representing inhomogeneities in the geographic space being modelled.” A set of potential factors was tested and the combination of factors that best predicted the land cover in the calibration and validation stages was selected. The set of potential factors included a random value, estimated travel times from the Kampala CBD and subcenters, the wetlands map, road density, flood depth and the additive inverse of the 2004 percentage of vegetation land cover.

¹The neighborhood of a cell in a grid of square cells consists of the adjacent cells to the one being considered (Benenson & Torrens, 2004); it can be defined by the cells immediately adjacent (first order), to those adjacent to the adjacent ones (second order), etc. The size of this neighborhood must be defined; see appendix C. In this model, the neighborhood is the average value of the built up percentage for the cells located in this spatial neighborhood.

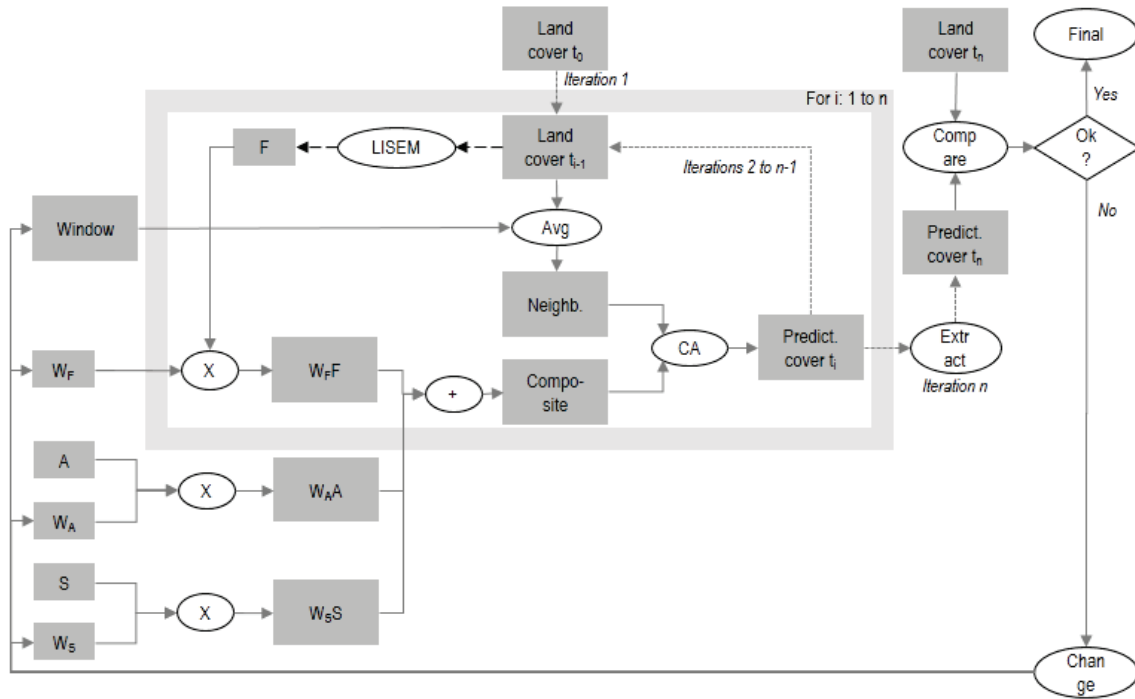


Figure 4.1: Schematization of the CA model: elements and calibration process for the full model

The transition rule, governing how the cells change within the simulation, does not follow the tradition of most CA models with a discrete state (in which the result of updating a cell is a change in the category of the automaton). Rather, the transition rule results in a gradual change analogous to the grey cell models of Yeh & Li (2001). In Yeh & Li (2001), the grey cell transition rule for an automaton, designated by its xy coordinates, is:

$$G_{xy,t+1} = G_{xy,t} + \Delta G_{xy,t} + r_e \quad (4.3)$$

and in this equation, the change $\Delta G_{xy,t}$ depends on both the neighborhood values and the composite synthesizing ancillary effects.

$$\Delta G_{xy,t} = f(Composite_{xy,t}, Neighb_{xy,t}) \quad (4.4)$$

r_e is a stochastic disturbance term, representing unknown errors, unaccounted factors and random decisions by agents. A modified version of the formulation by White & Engelen (1993)², as cited in Yeh & Li (2001), was used to quantify this term:

$$r_e = |[-\ln(\gamma)]^\alpha - 1| \quad (4.5)$$

with γ a random value between 0 and 1, and α a parameter to reflect the degree of dispersion in the simulation.

Unlike the CA model of Yeh & Li (2001), for the developed model, the state of Au is a vector of two dimensions, $\mathbf{G}$. To operationalize this, ΔG_t was simulated and used to estimate $G(1)_{t+1}$ as

²The modification was introduced to re-scale the values, such that (1) the minimum value is 0 and (2) the average value may be near the averages of other factors, especially the average built up land cover in the neighborhood of the cell. Additionally, possibly because of the computational algorithms embedded in it, the GIS software (ArcGIS 10.1) could not compute the original White-Engelen formulation.

$G(1)_t + \Delta G_t$; the percentage of bare soil, $G(2)_{t+1}$, was defined as a function of the fixed ratio of bare soil to built up areas in the initial period of the simulation. The rule is described specifically in section 4.2.

Finally, a global constraint, in the sense of White (1998), was introduced into the model because the total demand of built up area is thought to be exogenously determined by urban dynamics (e.g., immigration to Kampala, distribution of the population within the metropolitan region, economic development).

The final transition rule for $G(1)$ was specified as:

$$\Delta G_{i,t+1} = \begin{cases} SimDem_i, & \text{if } i \in [k, n] \\ 0, & \text{if } i \in [0, k] \end{cases} \quad (4.6)$$

such that:

$$400 \cdot \sum_{i=k}^n \Delta G_{i,t} \approx LD_{t \text{ to } t+1} \quad (4.7)$$

with $SimDem_i$ the simulated demand allotted to cell i , n the total number of cells in the study area and $LD_{t \text{ to } t+1}$ is the total demand for built up land in the period t to $t + 1$ (in m^2).

The index i denotes the suitability for development score of cell i . The only cells exhibiting land cover change (increase in built up land cover) are those which have (1) a suitability score higher than any non-developed cell and (2) all of which, jointly, sum up to the exogenous demand. The index i is a function of the $Composite_{i,t}$ and $Neighb_{i,t}$ values of each cell.

In theory, it is not necessary that each cell has a unique i value; it suffices that the variation in the index is enough to distinguish between the two groups of cells (i.e., updated and not updated). In practice, when factors with continuous variations - such as the neighborhood factor itself, the estimated travel times to the CBD and subcenters of Kampala, road density or non vegetation percentage - are included in the composite, i is transformed into a unique identifier. However, when the composite only includes factors with little variation, be it few values (such as the wetland factor, which has only three possible values) or an excess of cells with the same value (e.g., flood depth for which the majority of cells have a value of 0), the specified model is not efficient because the i value cannot separate with sufficient detail between the cells that should be updated and those that should not.

4.2 IMPLEMENTATION STRATEGY

The CA model developed has idealized space as a regular grid of square cells. Cell size was set at 20 m to conform with the data models of the calibrated openLISEM model³.

The calibration process was implemented in a step-wise manner, as described in the introduction. In the first step, a model based on a single factor was estimated for each one (only the model based

³The flood model was calibrated for the Upper Lubigi catchment with a resolution of 10 m; the results of this work, which were linked to land cover scenarios generated using statistical methods, were reported in Sliuzas, Flacke, & Jetten (2013). The reduction in spatial resolution is justified because the study area of this project is larger than the Upper Lubigi and, more importantly, because the software used for the CA model is much less efficient than the loose coupling of openLISEM and the statistical methods used in Sliuzas, Flacke, & Jetten (2013) (and in the linked work of Abebe, 2013 and of Fura, 2013).

solely on the neighborhood was a CA model proper), to understand the influence of each factor on the predictive character of the model. Based on these results, the factors that were expected to best predict the land cover were selected and tested, as well as alternative combinations. The final combination of factors was then calibrated assigning different weights and selecting the best, in terms of predictive accuracy.

Three issues must be resolved at each stage of the calibration process: (1) how much change into built up land cover can potentially occur in each cell, (2) which cells are more likely to change and (3) how many cells must change to fulfill the demand for land.

Issue (1) is tackled by exploring the potential determinants of land cover change for the Lubigi catchment (calibration data). Since no clear predictive relationships of the *amount* of development emerged⁴, an alternative simulation strategy was adopted: the amount of built up percentage increase was assumed to be spatially random - as is the data of Lubigi - but its logistic transformation was defined to have a normal distribution, with mean and standard deviation equal to the calibration data (i.e., $Mean_{LogisticTransf} = -1.2328$ and $s.d._{LogisticTransf} = 1.7605$).

Thus, the procedure to simulate the potential built up development percentage for each cell was: (a) each cell was assigned a random number between 0 and 1, taken to be the probability associated with the z value of the logistic transformation of urban development; (b) from the mentioned random value and using the mean and standard deviation, the simulated value of the logistic transformation was estimated; (c) using equation 4.8, the percentage of built up increase was estimated; (d) as a final verification, the sum of the built up increase for all cells was compared to the exogenous demand (the total built up increase for the calibration area during the period 2004-2010), in order to check that the former was larger.

$$SimDem_{xy} = \frac{e^{SimLog_{xy}}}{1 + e^{SimLog_{xy}}} \quad (4.8)$$

with $SimLog$ the estimated z value. Note that the logistic transformation was specified using natural logarithms, hence the use of e in the reverse transformation.

Issue (2) allows for the incorporation of spatial variation into the modeling framework. An allocation index, evaluating the spatial differentials in the potential for urban development, is defined as in equation 4.9:

$$i = \sum_k w_k f_{xy,k} \quad (4.9)$$

where w_k is the weight of factor k relative to the neighborhood effect, which is assigned a weight of 1 in order to compare the relative importance of other factors to it; and $f_{xy,k}$ is the value of factor k in cell xy .

Cells are sorted according to this i index. Development is allocated to (change in built up land cover is assumed to occur in) the set of cells with the highest allocation index. Formally, these would be the $n - k$ cells such that:

$$\sum_{i=k}^n \Delta G_{i,t} = LD_{t \text{ to } t+1} \quad (4.10)$$

$LD_{t \text{ to } t+1}$ being the total demand for built up land cover and i denoting each cell, characterized by its allocation index.

⁴However, the explanatory models did predict, with certain degree of success ($\hat{R}^2 \approx 0.30$), which cells change and which do not.

A limitation is introduced when allocating development in this all-or-nothing manner. For example, if two neighboring cells have development potential, i.e., predicted land cover change into built up, of 30% and 25%, within the model, the 30% cell will be completely developed before any change occurs in the 25% cell. Evidently, in reality, both cells may very well be partially developed simultaneously. The relatively fine grain of the CA space (the small size of the cells) should mitigate this problem.

Issue (3) is less straightforward than what is usual in raster-based models of urban growth because each cell contributes a different amount of development.

To select which cells do change in the simulation, fulfilling the constraint of equation 4.7, the following algorithm was implemented: (a) A *for* cycle, going from values 1 to m for parameter h , was defined. For each iteration: (b) The multiplicative inverse of the iteration ($1/h$) was estimated. (c) All cells such that $1 - i < h$ were selected. (d) For this selection, the sum of $SimDem$ was estimated. (e) This $\sum SimDem$ was compared to the exogenous demand, LD : if $LD > \sum SimDem$, all selected cells were considered as *CHANGE*, i.e., cells that need to be updated; if $LD < \sum SimDem$, all selected cells kept their condition from the *previous iteration*; in other words, if in a previous iteration a cell was assigned as a *CHANGE* cell, for all subsequent iterations it kept this condition. The algorithm ceases to add new cells into this *CHANGE* category when the restriction represented by equation 4.7 was met.

Since no clear empirical relationship was established between the land cover patterns of the bare soil and built up categories (see the discussion in appendix C), the assumption by Fura (2013) that the ratio of bare soil to built up land cover percentages remains constant in a cell, has been maintained. Two exceptions to this rule were considered: (a) that the sum of built up and bare soil land covers exceeds 100% and (b) that the percentage of built up land cover of 2004 was 0%, in which case the bare soil percentage was assumed to remain constant.

In the former case (overallocation), built up land cover is assumed to occupy all the space available, up to 85% of the total predicted land cover, the rest being allocated to bare soil. In this way, a fully occupied urban cell would be comprised of 85% built up area and 15% bare soil - mostly expected to be used as roads or walkways. The only exception to this rule is the following: if the 2004 built up land cover percentage exceeded 85%, it was kept constant for the second period (2010 for the calibration process). As for the bare soil percentage, it was estimated by multiplying the ratio of bare soil to built up from the initial period of the simulation, times the predicted built up percentage; if the sum of both estimates (of built up and bare soil) exceeds 100%, then the bare soil fraction was set at $100\% - \%Built\ Up$. Figure 4.2 presents a graphical representation of this process.

Further, prior to the calibration process, the bare soil was divided into two subcategories: the “on road” bare soil - the bare soil land cover which overlaps with the area of the road network - and the “off road” bare soil. Only the “on road” bare soil was updated; the “off road” bare soil was lumped with the vegetation land cover category (in the final stage of prediction, when urban growth had been completely assigned and the tarmac area had been accounted for, the remaining non developed area was divided proportionally - based on the percentages of the initial period - into vegetation and “off road” bare soil; this is a necessity if the result is to be inputted into the openLISEM model).

A final point of attention relates to the conceptualization of space. The basic framework, as stated, is a grid of square cells with a side of 20m. Each cell is an automaton, fully characterized by a set of seven percentages defining the area occupied by the seven land cover classes of the developed map (see section 2.2 and appendix A; the bare soil category, as stated, was divided into two sub-

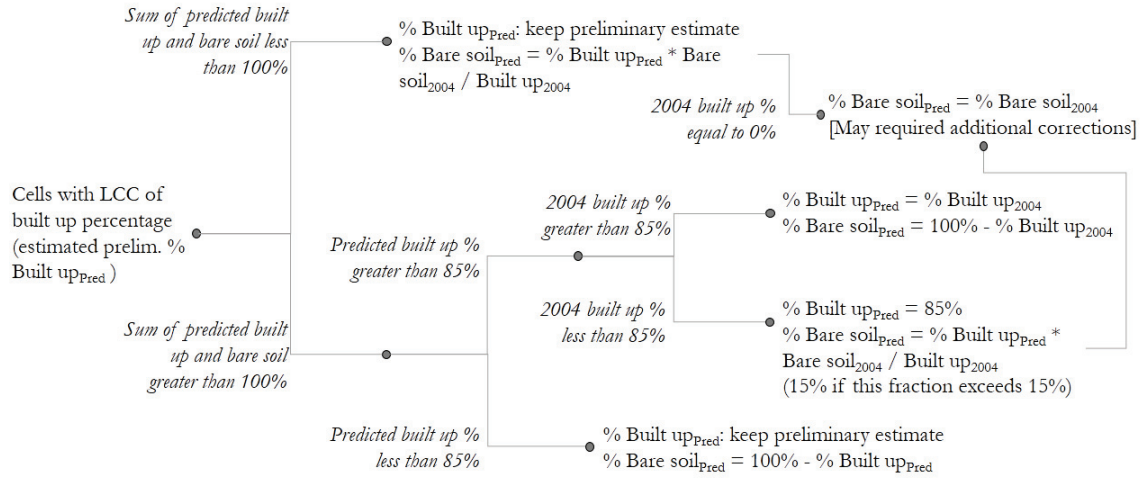


Figure 4.2: Rule for estimating bare soil percentage based on built land cover percentage in CA models

categories). Of these, two must be predicted by the CA model: built up land cover, the main objective, and the “on road” bare soil because its expansion is associated to urban development. These land cover categories may expand but they cannot, within the model framework, occupy the entire area of the cell. They can only substitute the vegetation land cover (and the “off road” bare soil) area. It is assumed that tarmac is essentially constant⁵, as is the water surface; the ‘no data’ category is excluded from all simulations due to the lack of information.

To incorporate this element into the model, the percentages of the non-dynamic categories are assumed as set. The dynamic percentages are converted by multiplying times a transformation factor, such that the percentages of the dynamic categories sum 100%:

$$Transf(\%LC_i) = \frac{100\%}{\%LC_{BU} + \%LC_{Rd\ BS} + \%LC_{Veg} + \%LC_{OffRd\ BS}} \cdot \%LC_i \quad (4.11)$$

where i denotes the categories of built up (BU), “on road” ($Rd\ BS$) and “off” ($OffRd\ BS$) bare soil and vegetation (Veg) land cover. This transformed percentage was used in the derivation of transition rules and the allocation of land cover classes. In both the calibration and simulation processes, the land cover percentages for the non-change land cover categories of the initial period were maintained constant. The remaining area was redistributed, according to the model’s transition rules.

⁵More precisely, that the changes in tarmac land cover cannot be properly predicted from the dynamics formalized in the model.

Chapter 5

Cellular Automata Model Calibration and Validation

5.1 PROCEDURE FOR CELLULAR AUTOMATA MODEL IMPLEMENTATION

The elements described in the preceding chapter (4) form the basis on which the model was implemented. Combining them, it allocates an exogenously determined demand for land using the following process:

1. The transformed percentages of built up, bare soil and vegetation land cover, for both 2004 and 2010, were estimated using transformation factors specific to each year and study area (Upper Lubigi for the calibration exercise, Nalukolongo for the validation section). The bare soil category was separated into “on road” and “off road” by (1) creating a buffer for the road lines, varying the width per type¹, (2) assigning as “on road” bare soil all the pixels which overlap with this buffer and as “off road” all other pixels.
2. The maps describing other factors were also estimated for each cell and the resulting averages were then normalized using linear maximum standardization criteria².
3. The transformed land cover percentages (estimated using equation 4.11) were calculated for all cells.
4. The simulated demand, *SimDem*, was calculated for all cells by assigning a random number, from 0 to 1, estimating the corresponding logistic transformation value (with the inverse normal distribution, and the mean and standard deviation reported in chapter 4) and applying equation 4.8 to it.
5. The potential future land covers for all cells, using transformed percentages, were estimated:
(1) The potential built up was estimated as the sum of the 2004 built up percentage and the simulated demand. (2) The “on road” bare soil percentage was estimated by multiplying this predicted built up percentage with the ratio $\%Rd\ Bare\ Soil_{2004}/\%Built\ Up_{2004}$ (both of these steps subject to the corrections outlined in figure 4.2). (3) The percentage of vegetation was estimated by calculating the percentage not occupied by built up and “on road” bare soil and multiplying it by the ratio $\%Vegetation_{2004}/(\%Vegetation_{2004} + \%OffRd\ Bare\ Soil)$. (4) Similarly, the percentage of “off road” bare soil was estimated by multiplying this same percentage by the ratio $\%OffRd\ Bare\ Soil/(\%Vegetation_{2004} + \%OffRd\ Bare\ Soil)$.

¹These widths were assigned from visual inspection of the imagery, as follows: A1 roads, 2m; footpath roads, 1.5m; main roads, 2m; minor roads, 3m; ROAD, 3.5m; railway, 0m; truck roads, 2m.

²If the factor favors urban development, (e.g., greater road density) is associated with more urbanization, the formula: $x = score/highest\ score$ was used. If the relationship is inverse (e.g., the smaller the travel time to the CBD, the greater the development potential), the formula: $x = (score - lowest\ score)/highest\ score$ was used.

6. The allocation index was estimated for selected combinations of factors and weights (which are documented in the following sections of this chapter).
7. The allocation index was used to identify the cells that must change to fulfill the simulation's restrictions (specifically, that the sum of potential development must be approximately equal to the exogenous demand, both for validation and calibration). For these cells, the potential land cover percentages were assigned. For all other cells, the 2004 land cover percentages were maintained.
8. All transformed percentages were converted into 'raw' percentages by applying the inverse transformation factor (equation 4.11) with the 2004 data.

The total demand allocated in the calibration process was: $LD_{t \text{ to } t+1} = 5378$. This figure results from estimating the differential between built up land cover in 2010 and 2004 for each cell in the Lubigi catchment, and summing up this percentage for all cells (if it were multiplied by $400m^2$, the resulting number would be the net area that changed into built up, in square meters; since the factor is constant, and therefore equally affects all cells, calculations have been performed without this multiplication). For the validation process, the assigned $LD_{t \text{ to } t+1}$ equaled 2470.

The allocation algorithm was run (1) dynamically, i.e., for each year sequentially and (2) with a precision (m value in the allocation algorithm) of 100. The yearly allocation proceeded as follows: since the period being simulated was 2004-2010, the exogenous demand was divided into six equal parts. For a single period, $1/6$ of the exogenous demand was allocated. Then, the resulting predicted land cover was used to re-calculate the neighborhood effect and, if the neighborhood effect was included in the allocation index, to recalculate it as well. In the following period, another $1/6$ of the exogenous demand was again allocated but using the updated index. This process was repeated six times - since there are six periods between 2004 and 2010. The object of this process is to identify cells that change. Implicit within the algorithm is that a cell will only change *once* within a single simulation; thus, cells that are identified as changing participate in later iterations with their updated value (in particular, the neighborhood average is estimated with the predicted value) but development is not allocated to them more than once (they are not considered as potentially developable once they have changed).

Eight factors have been explored as potentially contributing to explain change into built-up land cover during the 2004-2010 period in Kampala. Land cover change has been conceived as a process determined by the local environment (i.e., the immediate neighborhood of a given location), other variables at the same location - what Tobler (1979) termed *multivariate models*, history (land cover at that same location in preceding periods) and randomness.

As stated in the model design, to account for the random behavior of urban agents (developers, home and land owners, etc.) as well as unaccounted for errors and dynamics, a random term has been introduced into the model. Randomness should be especially important in the context of a city such as Kampala because of its high levels of informality, poor enforcement of existing regulations and generally little planning. These, coupled with relatively high supply of land, result in more dispersion (a weakening of possible agglomeration effects), which in turn reduce the explanatory power of the neighborhood percentage. This random factor was specified by assigning a random value, ranging from 0 to 1, to each cell and estimating the factor with equation 4.5. To complete this calculation, the factor α was set at 0.19 because the average value of all cells, if this factor is specified, is similar to the average value of the neighborhood percentage for 2004.

The "other variables" considered were related to regional accessibility (travel time to the CBD, travel time to city subcenters and road density), physical suitability (an index representing the

effect of wetlands and flood depth), and history (the additive inverse of the 2004 percentage of vegetation land cover).

The wetland impact was formalized thus: if the wetland is permanent, a value of 0 was assigned; for a seasonal wetland, these locations were coded as 0.5. All other locations were given a value equal to 1.

The flood depth was estimated using the openLISEM flood model calibrated for Kampala (Sliuzas, Flacke, & Jetten, 2013) but substituting the original land cover maps with those developed in this project. Flood depth was used as a possible criteria only for the Upper Lubigi (calibration) area because no soils data is available for the Nalukolongo (validation) area³.

As for the estimated travel times, their derivation was discussed in chapter 2 and the ArcGIS 10.1 ModelBuilder models used in the estimation, are reported in appendix B.

The non-vegetation percentage is considered as equivalent to a time lag: land cover would tend to maintain its character; thus, if the non-vegetation percentage (mostly built up, “on road” bare soil and tarmac) is relatively high, those cells should grow first - until they are filled up (until there is no more free space for development).

In general, for each factor, the effects of three possible weights were explored: 0.5, 1 and 2. In all cases, these weights are considered relative to the neighborhood factor - which is to say that the weight of this factor is 1 by definition and other factors represent half the effect of the neighborhood, equal to the effect of the neighborhood or double the effect of the neighborhood.

An allocation index was estimated for combinations of weights and factors. For each allocation index, a land cover prediction was computed by allocating the total demand to the most suitable cells (i.e., the cells with the highest value of allocation index). Subsequently, the resulting patterns were evaluated; the combination of factor and weight selection corresponding to the pattern that best approximates the 2010 land cover map was selected and used in each subsequent step of the calibration process. When introducing additional information did not result in a significant improvement, relative to the previous step of the calibration process, the factor was not included in the next steps.

5.1.1 Evaluation of predicted land cover patterns

The evaluation of a simulated land cover pattern, resulting from any step in the CA model calibration process, is accomplished by (1) computing a series of quantitative measures of the similarity between the 2010 land cover map and the prediction and (2) visually comparing maps of both data sets. The comparisons are performed on the percentage of built up land cover for 2010. They aim to assess the correspondence between the data model (the 2010 land cover map) and the model prediction, in three distinct ways:

- Average correspondence between prediction and map, per cell: It is quantified through two measures, a correlation coefficient (which is a biased measure, since the statistical distributions involved in the comparison are not normal) and a Kappa coefficient. The Kappa coefficient measures how much greater is the correspondence between two categorical maps than chance agreement; since categories are required, both maps were reclassified into five categories, using their built up land cover percentages (0-20%, 20-40%, 40-60%, 60-80% and

³This factor was explored using a preliminary model of flooding, the same reported in Mohnda (2013). The results of chapter 6 were estimated with an updated model, including new data from late 2013 and early 2014.

80-100%). The larger each parameter is, the greater the correspondence.

- Correspondence between the statistical distributions of the prediction and map: The Kolmogorov - Smirnov (KS) nonparametric test on the percentage of built up land cover was calculated, assuming two samples. The smaller the KS statistic, the closer the distributions are to each other (more precisely, the greater the probability that both samples were drawn from the same population).
- Overall spatial patterns: This aspect is explored by a combination of two measures: the estimation of overall global spatial autocorrelation, as measured by Moran's I (assuming first order adjacency between cells in a Moore neighborhood as the relationship), and the visual inspection of the maps. The parameter and map are separately estimated and compared; there is no formal quantification of their correspondence but the closer the results are among each other, the better the prediction.

5.2 EXPLORATION OF THE EFFECTS OF SINGLE FACTORS

As discussed previously, the first stage in the calibration involves an attempt to explain the 2010 land cover patterns for the calibration (Upper Lubigi) area using only a single factor as input in the model. Of the eight factors explored, only one yields what can be properly called a CA model - the model using the neighborhood factor as input, which is a strict CA model in the sense that the prediction depends solely on the neighborhood. The predicted land cover patterns are shown in figure 5.1; table 5.1 summarizes the statistical assessment of each prediction.

The statistics reported in table 5.1 are four: a Pearson correlation coefficient - the closer it is to one, the greater the correspondence between the land cover map and the prediction -, the Kolmogorov-Smirnov statistic, which measures the difference between two samples assumed to be taken from the same distribution (if the KS statistic is zero, the null hypothesis that both samples come from the same distribution cannot be rejected; thus, the smaller the KS statistic, the better the prediction), Cohen's Kappa, the product of reclassifying the built up percentage into five categories (and measuring how much better the prediction is compared to a random pattern; the higher this percentage, the better the prediction) and, finally, a clustering measure: the global Moran's I; for this statistic, the objective of each prediction is to replicate the value reported for the land cover map.

In addition, a fifth measure of goodness-of-fit has also been reported: the total allocated demand. The algorithm applied allocates the simulated demand until the condition $LD > SelDem$ (the exogenous demand to be assigned in each period is more than the demand being allocated in each iteration) is not met; when the condition is met, no further allocation occurs. If the allocation is accurate, the difference between these two values is small (by construction, the exogenous demand is always greater than the allocated demand). On the contrary, if the factor does not allow for a proper allocation, the differences between the target (exogenous) demand and the simulation is large - in table 5.1, which reports the total built up percentage for 2010, the difference between the 2010 land cover map and the sum of predicted built up percentage should be as small as possible.

An analysis of the statistics reported in table 5.1 does show some distinctive characteristics of each factor:

(1) The neighborhood and random factors both result in more dispersed predictions (lower Moran's I value), relative to the land cover map; travel times to CBD and subcenters as well as road density (the accessibility factors) produce an opposite effect, over-clustering (higher Moran's I value than the map). This dispersion of the neighborhood effect is consistent with a dynamic assignment,

based on predictions which are spatially random - thus creating an equivalent effect to the random factor.

(2) Two factors - wetland and flood depth - are not allocating correctly: the differences between the sum of built up percentage for their respective predictions differ in over 30% from the targeted (land cover map) sum. All other factors show differences below 10%, although the difference is surprisingly high for the road density factor (allocation accuracy should be a function of the diversity of the factor; wetland and flood depth are predominantly 0 value cell factors; this is not the case for road density, which is derived from a raster map that represents continuous variation).

(3) The best performing factors vary according to the statistic being analyzed. In terms of allocation and of the KS statistic, the best factor is the neighborhood effect, followed by the random factor. The flood factor shows the best correlation and the second highest Kappa value, possibly owing to the large amount of cells that maintain their 2004 value. Travel time to subcenters is a factor which results in the best value for Cohen's Kappa and the second best clustering level.

(4) The flood factor and the wetland factor show equivalent results. This is hardly surprising, since both maps are dominated by zero values and, in consequence, the allocation algorithm is likely selecting the same group of cells in both predictions. In fact, since for the first iteration $SelDem > LD$, it is constant and equal to the 2004 built up land cover.

Table 5.1: Predicted built up land cover percentage, 2010, using single factor models. Comparison of predictions and land cover map

Model	Correlation	KS	Kappa	Morans's I	Sum of BU%
Land cover map 2010	-	-	-	0.4517 $z = 237.8$	15950
Neighborhood factor	0.4519	0.0668 $Pr. < 0.001$	24.2%	0.2484 $z = 130.7$	15588
Random factor	0.4677	0.0744 $Pr. < 0.001$	24.9%	0.2260 $z = 118.9$	15260
Travel time to CBD	0.4570	0.0856 $Pr. < 0.001$	23.7%	0.4967 $z = 261.1$	15017
Travel time to subcenters	0.4905	0.0857 $Pr. < 0.001$	36.7%	0.4875 $z = 256.4$	14957
Wetland factor	0.6087	0.1904 $Pr. < 0.001$	30.1%	0.4055 $z = 213.5$	10614
Non vegetation (2004 LC)	0.5643	0.1179 $Pr. < 0.001$	27.9%	0.4193 $z = 220.5$	15105
Road density (2004 roads)	0.5057	0.1000 $Pr. < 0.001$	26.2%	0.5019 $z = 264.1$	14375
Flood depth (2004 LC, T = 2 years)	0.6087	0.1904 $Pr. < 0.001$	30.1%	0.4055 $z = 213.5$	10614

Correlation: Pearson Correlation Coefficient.

KS: Kolmogorov-Smirnov two-tailed test; null hypothesis: both samples come from the same probability distribution; all null hypothesis rejected with Prob. < 0.001 .

Kappa: Cells reclassified into five categories (0-20%, 20-40%, 40-60%, 60-80%, 80-100%); Cohen's Kappa between the land cover map and the prediction, was estimated for this reclassification, using the program of Reams (2010).

Moran's I: estimated using first order adjacency (including corners), row standardized

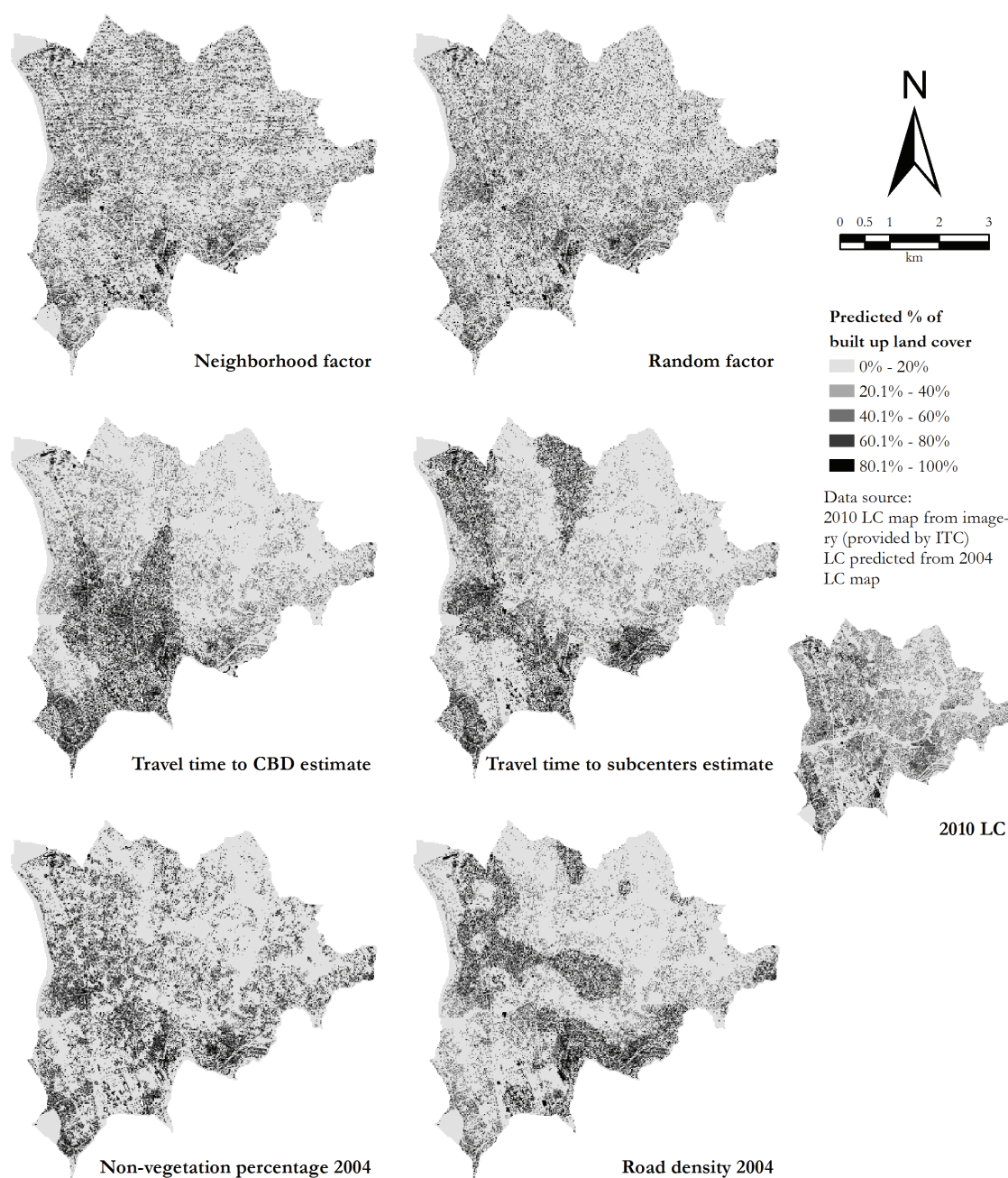


Figure 5.1: Land cover pattern predictions of the Upper Lubigi subcatchment, 2010. Single factor models

The patterns resulting from the predictions based on each factor - excluding wetland and flood factors, which resulted in biased allocations - show distinctive characteristics. Both the models based on the neighborhood and the random factor are more dispersed but the prediction based on the random factor is more uniform than the corresponding neighborhood factor prediction. The non vegetation factor model seems to replicate the 2004 built up land cover pattern (see figure 2.2), which interestingly also resembles, in broad terms, the built up land cover percentage pattern for 2010 (see the inset of figure 5.1; also reported in figure 2.2). The models dependent on accessibility

factors all show very distinctive patterns, with travel time to CBD predicting most of the development along the main roads and towards the southwest of the calibration area, the travel time to subcenters also showing a concentrated pattern but with several separate locations (along main roads) of development. Finally, the road density pattern predicts development along secondary and local road concentrations, rather than the main roads.

Synthesizing, it is evident that no single factor is sufficient to obtain an accurate prediction of the 2010 land cover. The final model must include information of several factors. As a preliminary model, four factors are selected: (1) the neighborhood effect, both because of its potential according to the KS statistic and because of theoretical reasons (CA models have been found to be very successful in predicting land cover patterns, as in Vermeiren et al. (2012); Abebe (2013); Fura (2013)), (2) the non-vegetation percentage, functioning as what Tobler called a historical model (i.e., the equivalent of a time lag in econometric models), (3) travel time to subcenters due to its success in terms of statistical measures (Moran's I and Cohen's Kappa), and, finally, (4) the wetland factor because it is very evident, from the sharp edges that can be seen in the inset of figure 5.1 (especially upstream), that the area occupied by wetlands is less developed than surrounding locations. This model, as well as alternative formulations, is developed in section 5.3.

5.3 SELECTION OF FACTORS COMBINATION AND WEIGHTS

A model with the four factors selected in the preceding section was estimated and its resulting predictions are assessed through table 5.2 and figure 5.2. Two variations were also included. In one version, the wetlands map was substituted by the flood depth as a restriction. Recall from the discussion in section that the characteristics of these single factor models make their explaining factors essentially equivalent; additionally, their spatial patterns are also similar (i.e., flooding occurs mostly in areas designated as wetlands by this map, even if some of these locations have been invaded by other land covers) - if anything, the wetlands area is larger than the flooded area. The second alternative version substitutes the optimal accessibility factor (travel time to subcenters) with a more standard variable (travel time to CBD) because the latter has been much more widely used in applications and it also plays a much more prominent role in theoretical models explaining the location of urban activities.

Four out of the five criteria shown in table 5.2 show very little difference in terms of the quality: in both correlation and Cohen's Kappa, four out of the five models perform equally well (the auxiliary model 2, evaluating the effect of the neighborhood factor and the travel time to subcenters is the only clearly inferior one); the Komogorov-Smirnov statistic mirrors these results but in reverse, model 2 is the best and all others are more or less equivalent. Clustering levels are also broadly equal except for the simpler (i.e., including only two factors) auxiliary models 1 and 2, which perform worse. Models 1 to 4 all show final allocation within 7% of the target, model 5 has the highest error (marginally over 10% difference) followed by model 6 (marginally under 10% difference). Note that this latter performance, while borderline acceptable, should not disqualify the models since increasing precision on the allocation is a matter of adjusting the model's parameters (in these evaluations, it has been kept constant because of computability and for comparison purposes).

The patterns shown in figure 5.2 do exhibit noteworthy differences.

(1) All models present a pattern that is more clustered than the land cover map. This problem is less acute for model 1 and more serious in model 2 - suggesting that the non-vegetation factor has the effect of dispersing development towards the central-eastern parts of the study area whereas

the accessibility factors introduce an opposite influence (clustering towards the generally more accessible west).

(2) All models over-allocate development in the northwestern corner - and more generally in the northern part - of the study area. In this sense, model 6 produces the best prediction because it seems to allocate more evenly than other models.

(3) What appears to be a strong restriction along wetland land cover (in the land cover model, inset of figure 5.2) is being well captured in the predictions when the wetlands map is incorporated; but increasing its importance may result in better predictions.

Because model 1 is too simple and an accessibility component is required, both theoretically and to evenly allocate development, the use of model 1 was not considered. Instead, model 6 was selected (and despite not being an optimal factor, travel time to CBD was chosen over travel time to subcenters).

The preceding analysis suggests that: (1) a weaker influence of travel time to CBD might improve the pattern in that it would be less clustered (it should promote more development in the north-eastern corner of the calibration area) and (2) a stronger wetland factor might contribute to steer development away from unsuitable areas, more sharply showing the restriction evident in the land cover map along the creek (see inset of figure 5.2).

Additional models using weight factors of 0.5 to weaken clustering effect of travel time to CBD

Table 5.2: Predicted built up land cover percentage, 2010, using multiple factor models with equal weights. Comparison of predictions and land cover map

Model	Correlation	KS	Kappa	Morans's I	Sum of BU%
Land cover map 2010	-	-	-	0.4517 $z = 237.8$	15950
1. Neighb. + Non-vegetation factors	0.5526 $Pr. < 0.001$	0.1103	27.7%	0.3938 $z = 207.1$	15017
2. Neighb. + TT to subcenters factors	0.4825 $Pr. < 0.001$	0.0754	25.3%	0.3828 $z = 201.3$	15273
3. Three factors	0.5517 $Pr. < 0.001$	0.1105	27.7%	0.4314 $z = 226.8$	14908
4. Three factors + wetland	0.5476 $Pr. < 0.001$	0.1084	27.8%	0.4379 $z = 230.3$	14896
5. Three factors + flood depth	0.5566 $Pr. < 0.001$	0.1216	28.3%	0.4318 $z = 227.0$	14307
6. Neighb. + non-veg. + TT to CBD + wetland	0.5507 $Pr. < 0.001$	0.1200	27.9%	0.4326 $z = 227.5$	14420

Three factors: neighborhood factor, non-vegetation percentage (2004) and estimated travel time to subcenters

Correlation: Pearson Correlation Coefficient.

KS: Kolmogorov-Smirnov two-tailed test; null hypothesis: both samples come from the same probability distribution; all null hypothesis rejected with Prob. < 0.001 .

Kappa: Cells reclassified into five categories (0-20%, 20-40%, 40-60%, 60-80%, 80-100%); Cohen's Kappa between the land cover map and the prediction, was estimated for this reclassification, using the program of Reams (2010).

Moran's I: estimated using first order adjacency (including corners), row standardized

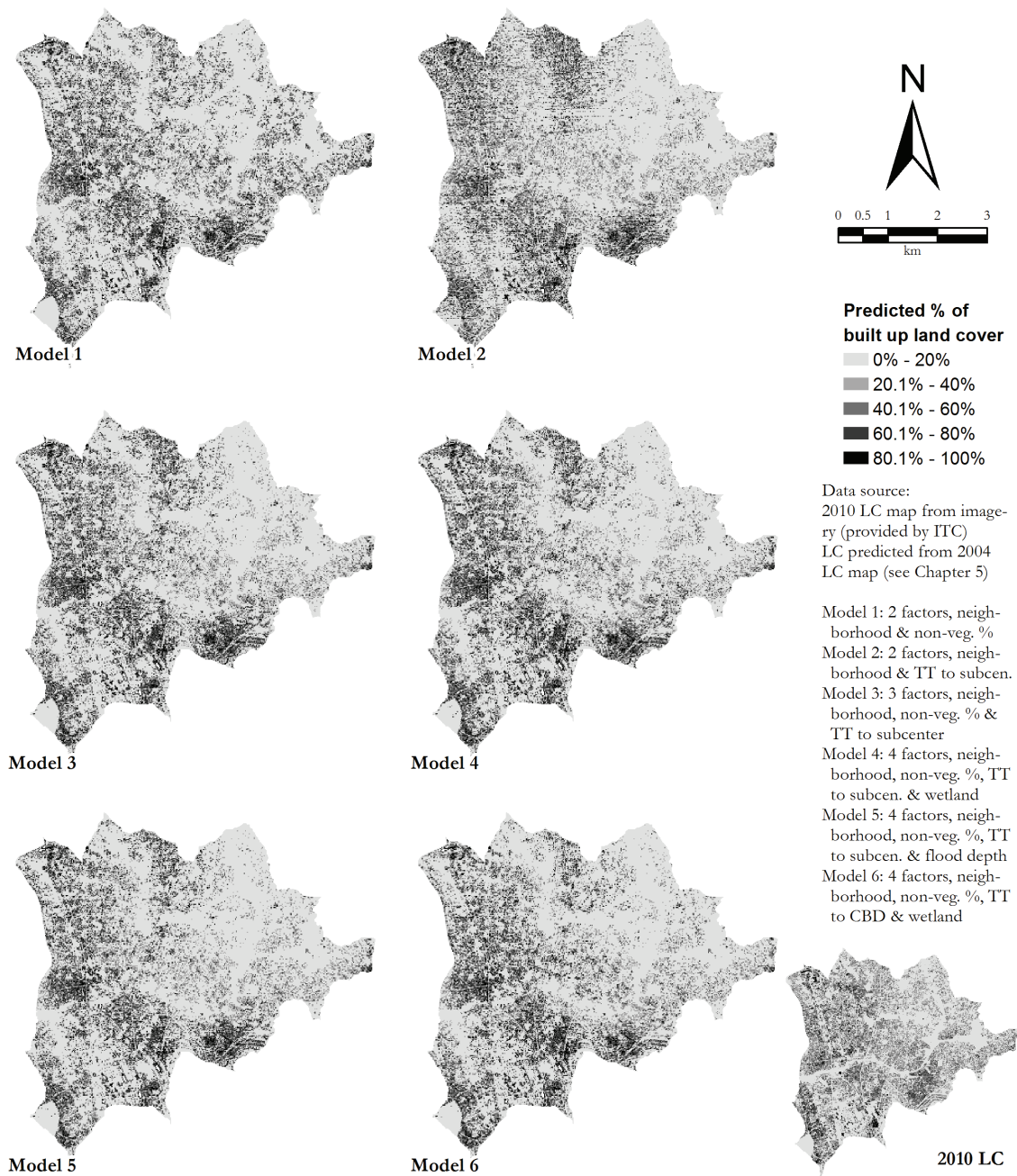


Figure 5.2: Land cover pattern predictions of the Upper Lubigi subcatchment, 2010. Multiple factor with equal weights models

and non-vegetation percentage factors and of 2 to strengthen the wetland boundaries have been calculated. The predictions of the combinations estimated, which are in line with the heuristic discussion of the previous paragraph, are reported in figure 5.3 (statistical analysis of the outputs was also undertaken but, since no clear differences were found, it is not reported).

Of the patterns resulting from the formulated models, model 4 (wetland factor weight of 2, and travel time to CBD and non-vegetation percentage weights of 0.5) was judged to be the best: while

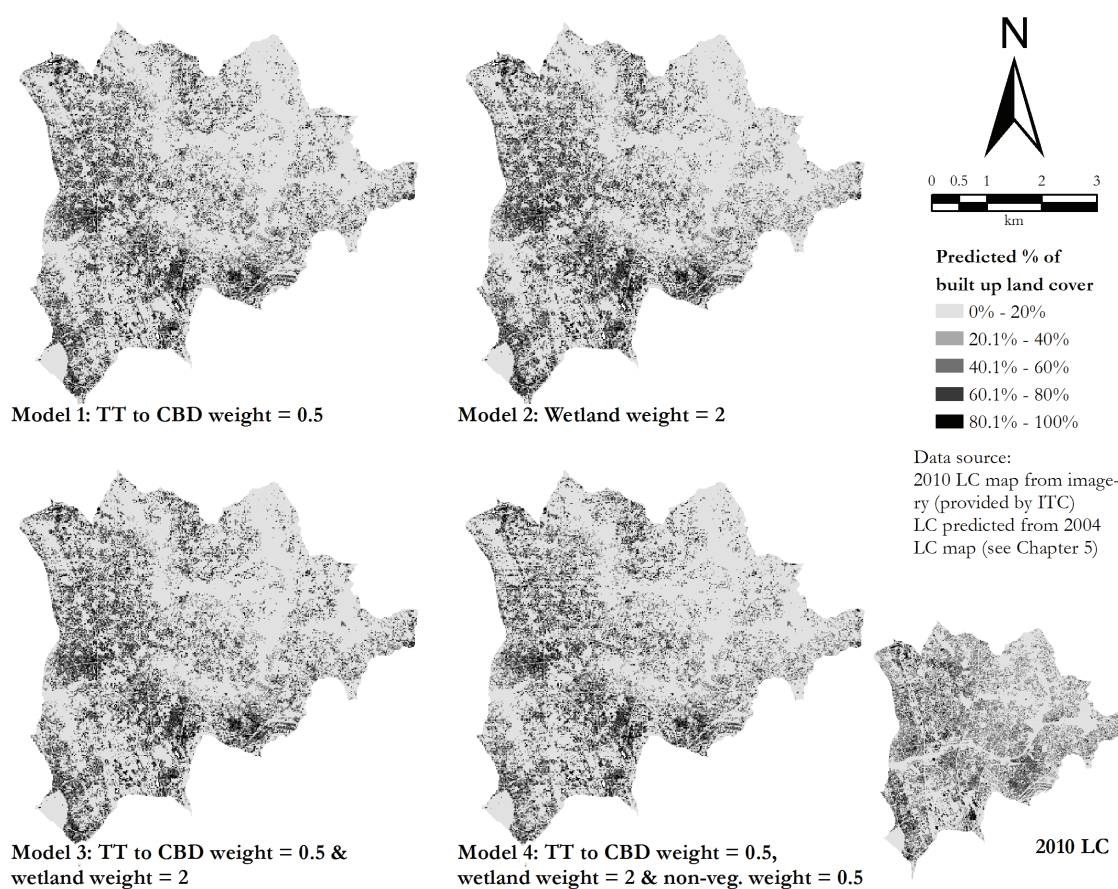


Figure 5.3: Land cover pattern predictions of the Upper Lubigi subcatchment, 2010. Multiple factor with varying weights models

very similar to models 1 and 3, it is less clustered in the western part (this does not imply more development in the central and eastern parts of the study area but it does resemble the land cover more than the alternatives). The weight specification, though, does not seem to have too large an influence on the predicted land cover patterns - it is certainly much less than the incorporation or exclusion of any given factor.

5.4 USE OF LAND USE DATA TO REFINE CALIBRATION RESULTS

The selected model factors and weights generally provide adequate levels of goodness of fit for the predicted patterns of built up land cover. The main limitation is the model's inability to reflect increased dispersal of development in the central and eastern parts of the study area.

It is possible that this limitation is partially explained to the neighborhood effect and the non-vegetation percentage of 2004, both of which assume development near existing infrastructure, regardless of the infrastructure's land use. However, this assumption is not completely proper because some of the existing built up land cover is occupied by institutional land uses in areas (outlined in figure 5.4) where, because of this, development should not be expected. This is particularly the case of the Makerere University.

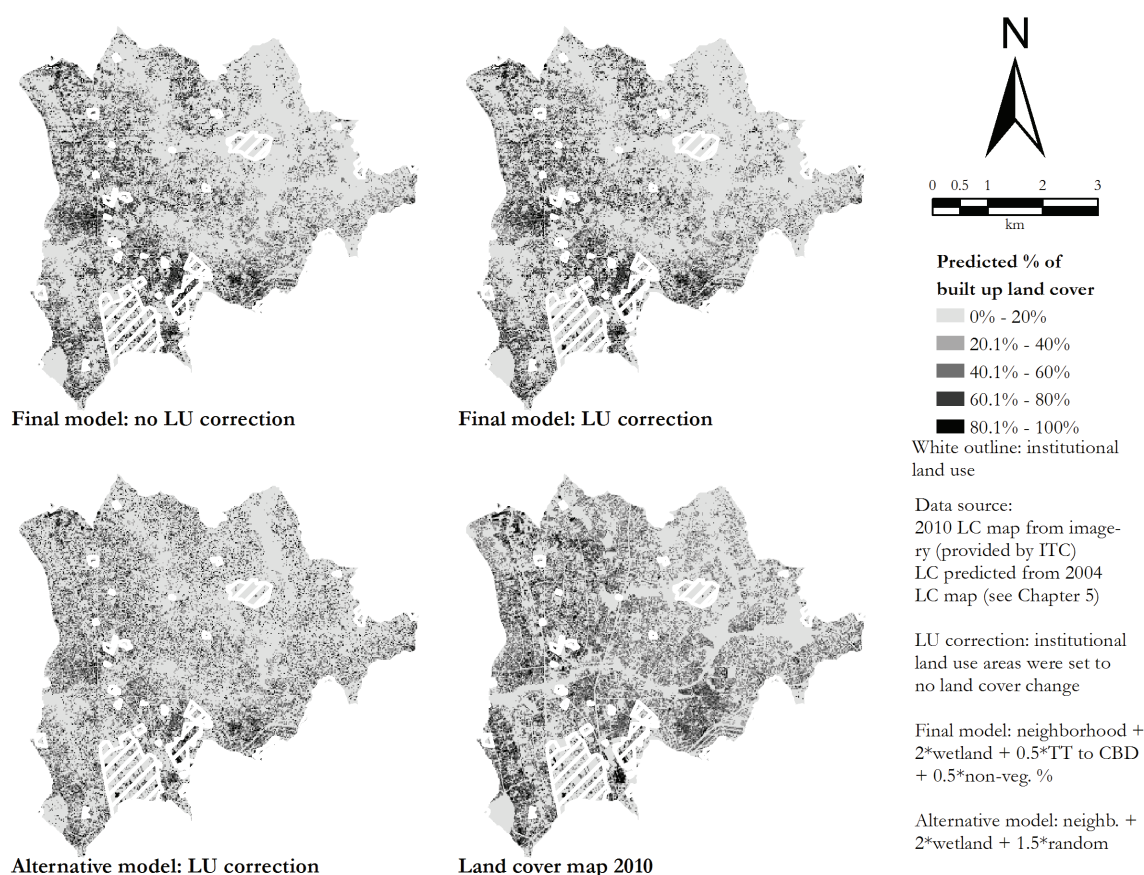


Figure 5.4: Land cover pattern predictions of the Upper Lubigi subcatchment, 2010. Final model and land use restrictions

To incorporate this element into the simulation, all cells having their centroids within these institutional land uses were selected and the potential land cover change was set to be equal to the 2004 land cover percentages⁴. In other words, for these cells, there is no possible land cover change by construction but they do maintain an attraction effect (they contribute to the neighborhood effect of their neighborhoods). This is consistent with expected dynamics: locations near these institutional land uses would be very attractive for development.

As can be seen by comparing the results of applying this specification, the differences introduced by the correction are not very large. The main difference is, perhaps, a better match between the land cover map and the restricted version of the final land cover model in the Makerere University campus on the southwestern corner of the calibration area.

Since the predicted pattern is still too clustered (relative to the land cover map), an alternative model was heuristically defined to generate the opposite pattern, i.e., one which is dispersed. As can be seen in figure 5.4, it included only the neighborhood factor and the wetland factor, and in addition to them, the random factor. The result, predictably, was an excessively dispersed pattern. What is remarkable, though, is that the dynamic CA component - despite the fact that most cells

⁴The areas excluded from development correspond to what is termed as institutional in the major land use category of the 2002 land use map of Kampala. The data was provided by ITC.

do not change during the simulation - is not sufficient to maintain the agglomeration effect. It is also remarkable that, for all predictions (but especially for the alternative model), the area occupied by streets is more weakly marked than in the land pattern. However, it is difficult to incorporate this element, since a dynamic updating of street expansion is a very complex problem in itself.

The result of this calibration section is the selection of a final model. This model is comprised by (1) the algorithms for allocating demand: selecting the cells that should change and dynamically updating the neighborhood characteristics of the land cover map in each intermediate step (see Appendix D for schematic views of the ArcGIS 10.1 ModelBuilder models used and a brief explanation of them) and (2) the selection of parameters to be set in these algorithms in order to obtain the best possible prediction of the future land cover. This second part involves the selection of the factors that are to be used in the allocation process (the neighborhood effect, the 2004 non-vegetation land cover, the wetlands map and the estimated travel time to the CBD), the weight of each factor relative to the neighborhood effect (for the wetland factor, 2, for the travel time to CBD and non-vegetation land cover, 0.5) and the introduction of a restriction based on land use, excluding allocation on institutional land uses. This final element does not result in an improvement on the predicted land cover patterns but it is theoretically and empirically sound and, in this sense, it represents an improvement on the model; for this reason, it was kept.

5.5 CELLULAR AUTOMATA MODEL VALIDATION

The final version of the calibrated model was used, jointly with the 2004 land cover map of Nalukolongo, to predict the 2010 land cover of this subcatchment. The resulting pattern is shown in figure 5.5.

Comparing the patterns shown in figure 5.5, it can be generally seen that the model was successful in reproducing the dynamics leading to them. Some specific successes of the CA mode include its ability to reproduce the restrictions imposed by institutional land uses - in the case of Nalukolongo, as represented by the absence of development in two large areas of the eastern border. But, also, the same limitations discussed for the calibration models exist.

In particular, the validation model's prediction is more clustered than the land cover map of 2010. This prediction *does* capture some of the dispersed development in the southwest of Nalukolongo. However, this should be attributed to the better performance of the Nalukolongo 2004 land cover map in identifying scattered development (which, in turn, is explained because of greater contrast between buildings and their surrounding vegetation in Nalukolongo vs. buildings mostly surrounded by bare soil in the more developed Upper Lubigi) rather than the predictive process itself (see figure 2.2 for the land cover patterns of both areas).

A second limitation has been detected with the validation results: the prediction does not account properly for the street pattern. This is evidenced by the more disorderly predicted pattern of the central-north and western-north areas of Nalukolongo, relative to the land cover map. Again, it is likely that the explanation can be found in the initial input (the 2004 land cover map): in the 2004 image, the differentiation of tarmac roads was much more challenging than in its 2010 equivalent, because of greater confusion with bare soil.

The statistical assessment of the simulated 2010 built up land cover pattern for Nalukolongo (validation area) is shown in table 5.3. It is similar to the previously discussed results for the calibration stage; if anything, results for Nalukolongo are slightly more accurate than the final calibration results. The correlation and Kappa coefficient are larger, the KS Statistic smaller and the difference

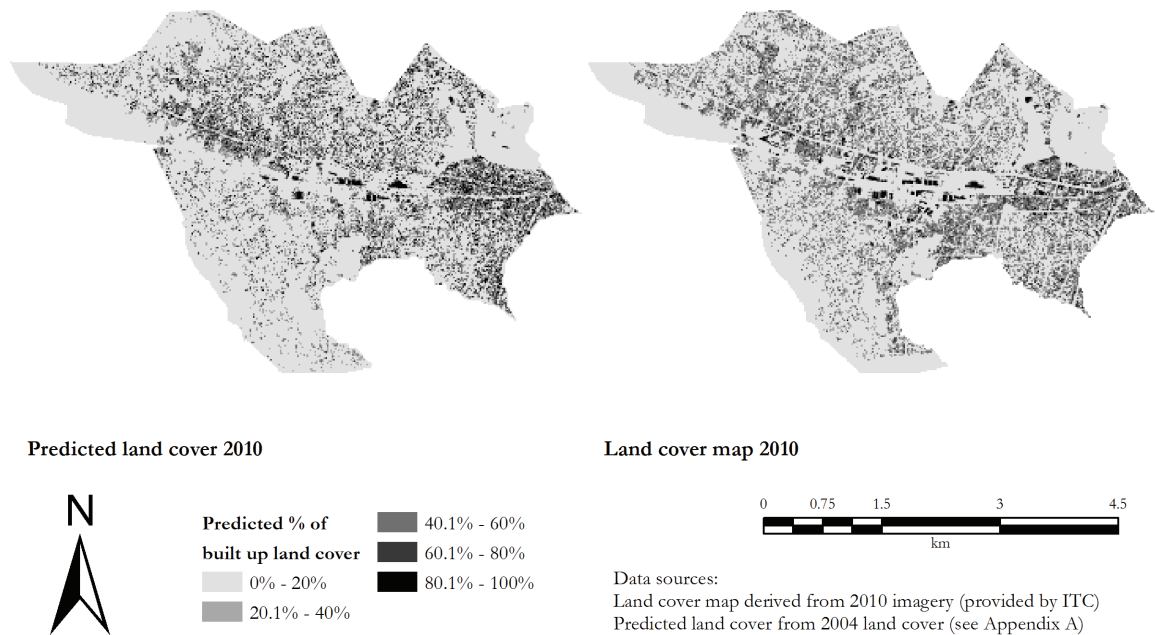


Figure 5.5: Land cover pattern predictions of the Nalukolongo subcatchment, 2010. Final model and land use restrictions

between the land cover map's and the prediction's Moran's I is less, for Nalukolongo than for Upper Lubigi. Possibly, this may be a consequence of Nalukolongo being a smaller and less complex area.

Table 5.3: Predicted built up land cover percentage, 2010, of Nalukolongo using the final calibrated model. Comparison of predictions and land cover map

Measure	Model Pred.	Land cover map
Correlation	0.6003	
KS Statistic	0.0531	
	$Pr. < 0.001$	
Kappa Coeff.	31.8 %	
Moran's I	0.4200	0.4462
	$z = 167.5$	$z = 175.7$
Sum of built up %	7426.0	7688.2

Correlation: Pearson Correlation Coefficient.

KS: Kolmogorov-Smirnov two-tailed test; null hypothesis: both samples come from the same probability distribution; all null hypothesis rejected with Prob. < 0.001.

Kappa: Cells reclassified into five categories (0-20%, 20-40%, 40-60%, 60-80%, 80-100%); Cohen's Kappa between the land cover map and the prediction, was estimated for this reclassification, using the program of Reams (2010).

Moran's I: estimated using first order adjacency (including corners), row standardized

In synthesis, the validation process can lead to three main conclusions: (1) The specified CA model is performing well, in terms of predicting the general spatial pattern of built up land cover. (2) The results are very sensitive to the input data, perhaps even to a greater extent than to the factors selected for prediction (although further analysis is required in this regard). (3) Statistical results support the comparison of visual patterns: the quality of predictions for the validation area is similar to those of the validation area (even slightly better).

5.6 DISCUSSION OF CALIBRATION PROCESS: METHODOLOGICAL RESULTS

The developed model has been, overall, very successful in replicating the 2004-2010 urban growth process. The predicted patterns are realistic and, generally, the impact of factors corresponds to theoretical expectations on their relative importance. The flexibility implicit in the definition of the allocation index (transition rules), their simplicity and the possibility of introducing restrictions (e.g., the discussed institutional land use restriction) all expand the possibilities of scenario simulation, relative to spatial statistical models previously calibrated for Kampala.

However, limitations do remain: the resulting patterns are (still) more clustered than the actual resulting map⁵. Better predictions may be achieved by revisiting the introduction of randomness into the model.

The second main limitation refers to the processes included: simulated potential land cover change assumes an overall process of urban growth. It does not reflect variations within this process - e.g., demolition of structures - nor other, non urban, land cover transitions. The model is effective within problems for which the main land cover disturbance is due to urban expansion within an urban area. It is less likely to succeed in other contexts (or, at the very least, it would require substantial changes, especially in the simulation of potential land covers). More subtly, all land cover changes are consistent with a single driver (i.e., all potential land cover changes are derived from the single initial change in built up percentage); in situations where two land dynamics coexist independently, the model will also require substantial adjustment of the allocation procedure, since it assumes only one demand for land (that of built up) controlling all change.

Interestingly, as was described on section 5.3, the modeling results were more sensible to the amount of information included in the transition rules (allocation index) rather than the different combinations of these factors. In general, including four factor maps produces better predictions than including merely two (of course, some specific exceptions do emerge, especially on statistical indices, but visual assessment of patterns jointly interpreted with all possible combinations does suggest strongly that more information leads to better results). Further, the final model specification is robust to different combinations of weights (once the factors have been selected, little variation is introduced by the weights; unreported preliminary tests with a wider range of weights combinations established this). However, it is also clear from the preceding analysis that information should not be blindly included.

Section 5.2 described different effects of each factor on the land cover pattern; but it also allowed to group these factors into categories having the same effect (e.g., both the random factor and the neighborhood effect disperse development, all accessibility maps lead to increased clustering, etc.) Two factors from the same group introduce only confusion but, also, factors with opposite effects, if not supported by substantive reasoning, may simply cancel each other out. Care should be taken in selecting not only the amount but also the specific factors to be included; while this process

⁵Because of how they are specified, this problem is more acute for spatial statistical models, such as those in Fura (2013) and Vermeiren et al. (2012)

Table 5.4: Pearson Correlation between factors and resulting 2010 built up land cover percentages

Factor	Upper Lubigi		Nalukolongo	
	Model pred.	2010 LC map	Model pred.	2010 LC map
Full area				
Neighborhood factor	0.4679	0.1992	0.4393	0.2344
Non vegetation (2004 LC)	0.5642	0.5561	0.5455	0.5396
Travel time to CBD	0.2445	0.2228	0.3072	0.2420
Wetland factor	0.1110	0.0593	0.1669	0.1278
Weighed summation	0.5248	0.2696	0.5785	0.3923
Change cells				
Neighborhood factor	-0.0931	-0.2380	-0.1803	-0.2821
Non vegetation (2004 LC)	-0.0338	0.1818	0.1789	0.3328
Travel time to CBD	0.2036	0.2492	0.1349	0.1343
Wetland factor	–	–	–	–
Weighed summation	-0.0279	0.0041	0.0444	0.0787

Full area correlations include all cells within the study area; *change cells* only includes cells that were predicted to have been changed in the model simulation.

is mostly heuristic, an initial assessment of each factor separately (and in particular, a statistical assessment of the resulting pattern, which describes the relationship between the factor and the land cover pattern) contributes to the overall understanding of the relationships involved and how the final pattern emerges from their interaction.

The assessment of preliminary versions of the model did not support the inclusion of the random factor, despite the initial notion of its importance in the context of Kampala. This is likely due to the large amount of randomness already present in the simulated demand. Recall that the potential land cover change into built up, while following a normal distribution, is spatially random. This implicit randomness is compounded by the dynamic updating of the neighborhood factor (its estimation based on the updated data, i.e., with some cells having changed but not all, thus introducing different levels of randomness in each iteration). In conclusion, the neighborhood effect and the random factor seem to have equivalent effects, with the neighborhood being more appropriate for the prediction of land cover patterns.

The analysis of how much each factor contributes to explain the simulated patterns is of great interest. Simulated patterns (summarized as the percentage of built up land cover) and their correlations with each factor are shown table 5.4. These correlations are, in general, consistent with theoretical expectations for all factors: non vegetation land cover (the 2004 pattern) most strongly correlates with both the prediction and the land cover map, followed by the neighborhood effect, travel time to CBD and the wetlands factor. No wetland area attracted any development, displaying this factor's characterization of a strong restriction of suitability on development.

As should also be expected, correlations between cells changing with the pattern are much lower than with the overall map - and indeed much less systematic. The general pattern has a very strong inertia, represented by the non vegetation factor as well as the neighborhood effect (to a certain extent), which should account for this difference. The magnitude of the difference may be considered too large; however, visual assessment of the patterns (figures 5.1, 5.2, 5.3, 5.4 and 5.5) belied this apparent mismatch.

The apparently odd result in table 5.4 is the very poor correlation between predicted built up percentage and the neighborhood effect for cells that change. The neighborhood was dynamically calculated but the neighborhood matched to each percentage of built up land cover was the value of assigned to the period in which change occurred. The importance of the neighborhood factor was made apparent by early, unreported, versions of the model. These early developments allocated for the entire 2004-2010 period based on a single, static neighborhood effect. When the neighborhood effect was included as a dynamic element, the improvement on the predicted land cover patterns was very clear in a simple visual comparison of both maps.

The negative correlation shown in table 5.4 between the neighborhood effect for change cells and the built up land cover prediction is due to a different relationship: the amount of change, as has been stated, was set to be spatially random, and in consequence, randomly related to the factor values that explain change. But the resulting built up land cover percentage is also capped; the sum of the base year plus the assigned demand cannot exceed 85% (except if in the base year it does, in which case it cannot change by definition). This implies that cells can be divided into two groups, those likely to change and those likely to remain constant, the latter including consistently high values of built up land cover percentages.

Looking to the use of the CA model for simulation, it is very important to underline a feature of the allocation procedure. Allocation proceeds by identifying cells that change within the simulation. These changed cells are assumed to transform into a potential land cover change, in turn simulated statistically. The statistical simulation was based on the mean and standard deviation derived from analyzing the 2004-2020 built up percentage increase. Therefore, simulated potential land covers are consistent with a growth process within this time scale. There is a strong assumption that a cell will only develop once within such a time frame; with longer periods being studied, this assumption no longer holds - and the longer the period to be simulated, the less likely a cell is to change (exhibit development) only once. Thus, when simulating a ten year period, it is more proper to sequentially simulate to five year periods than a single allocation of the 10-year expected demand for built up land cover growth.

Chapter 6

Scenario Development: Simulating Population Growth and Land Policy Interventions

Because the data models, the different dynamic processes and their interactions are all subject to considerable uncertainty, to derive specific information of policy options a scenario development approach has been adopted. Further, diverse scales of time and space are involved (e.g., of population growth, urban development, flooding). This compounds the uncertainty problems.

A set of scenarios was proposed for the Upper Lubigi subcatchment, with the aim of exploring the consequences of a range of conditions impacting flooding and land cover patterns. These scenarios were projected for a 10 year period (adopting as the base year 2010¹).

Each scenario has been designed as a specific combination of demographic and policy variables. They have all been projected with a variant of the calibrated model - adopting the same factors and most weights (with the exception of the wetland weight, which was reduced from 2 to 1 in order to reflect the fact that, as pressure for development increases, the restrictions posed by unsuitable land will decrease). Figure 6.1 schematizes this logic.

As can be seen, two main components determine each scenario. The land use distribution encompasses two domains: (1) the factors controlling where will land cover change occur - essentially the calibrated model - and (2) the policy objectives. These are introduced into the synthetic scenario through a geodesign approach, for example specifying a rule to cap the amount of development in each cell. Population growth is seen as an exogenous disturbance of the system and, as such, it is assumed to determine the entire demand for built up land cover - not only for residential uses but also for other urban functions (this is mostly true for commercial uses since their clients are residents but it is also, to a lesser extent, for other functions, more through co-location than a substantive dynamic proper).

These two components are linked: population is transformed into area demand by a density factor (which is also a policy variable, in the sense that it controls the target density of development); this area is then allocated using the derived CA model. If restrictions are imposed, the corresponding cells are retired from development by selecting the cells subject to the restriction, changing the initial land cover and setting it to be equal to the predicted land cover (land cover will not change for these cells in much the same manner as the institutional land use restriction already incorporated into the base year). Further, if this change involves the demolition of existing built up land cover (e.g., as in the evacuation of currently flooded areas), the demolished area was added to the exogenous land demand.

¹While 2010 land cover was defined as the base condition, the drainage system introduced into the model reflects the conditions of 2013 (especially, improvements in the main channel). This is because such improvements have already been implemented and because infrastructure is assumed as a constant determinant within the simulation process, i.e., it is not of interest to evaluate differentials introduced in impacts by this infrastructure improvement.

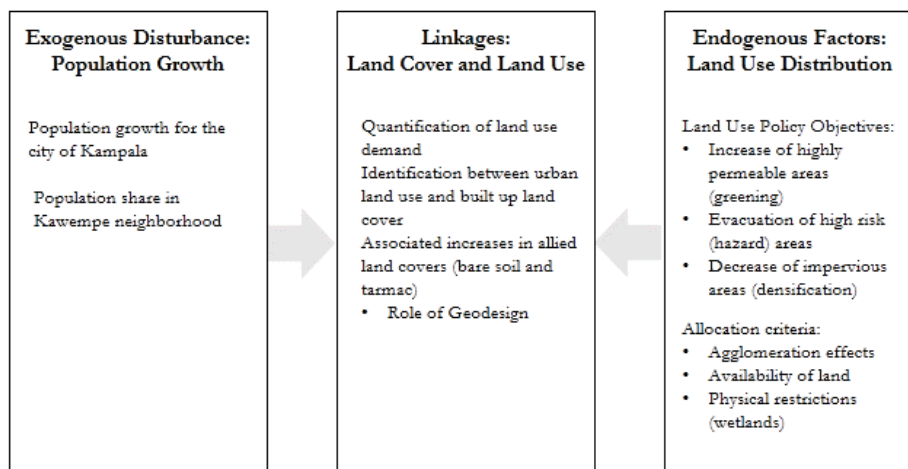


Figure 6.1: Scenario Development Logic

All land use policy formalizations are specifically described in the following section.

Since the area of the Upper Lubigi subcatchment does not correspond to any administrative unit for which population data is available, the population of the Kawempe division was adopted as equal to that of Upper Lubigi. It is a fair approximation, since 69.3% of the Kawempe division's area (total area of Kawempe: 3089 ha) is within the Upper Lubigi catchment - and the area that is not superimposed is the northern extreme of Kawempe, located the farthest from Kampala's CBD. Similarly, 76.3% of Upper Lubigi's area (total area of the subcatchment: 2805 ha) is within the Kawempe division. The scenarios assume the population share of Kawempe, relative to the entire city, to be constant in time. In other words, migration between parts of the city was not considered as a variable in the scenario design. A much more careful study of Kampala's spatial demographic dynamics is required before such an element is introduced in the simulation process (although it would be easy to implement, simply by introducing varying shares of population for Kawempe).

6.1 SCENARIO SPECIFICATION

6.1.1 Demographic Scenarios

Demographic data inputs have been defined based on the available population data. Trends were analyzed from data of the United Nations Statistical Unit. As can be seen in figure 6.2, three possible trajectories were considered, that of the projections adopted by the UN (4.8%) and two additional ones: a high growth trajectory, for which the interannual growth rate of the UN was set to 6.4% (a 1/3 increase from the trend condition), and a low growth rate (2.4%, a halving the UN's rate). All population estimates are for the entire city of Kampala and were calculated using as a base the year 2000 and applying the compound interest formula.

A cursory analysis of the trajectories clearly reveals that the high growth rate does not result in a

realistic forecasts, as it is unlikely that Kampala will quadruple between 2000 and 2020 (although it must be noted that the proposed Kampala Physical Development Plan uses projections that, by 2040, estimate the city's population at over 10 million residents). The high growth scenario is useful to explore scenarios in which Upper Lubigi is fully developed (most of the simulated land cover demand is materialized, i.e., transformed into built up land cover by 2020). The low growth hypothesis informs on potential costs of development measures: is the implementation of a policy justified, even if growth is lower than projected?

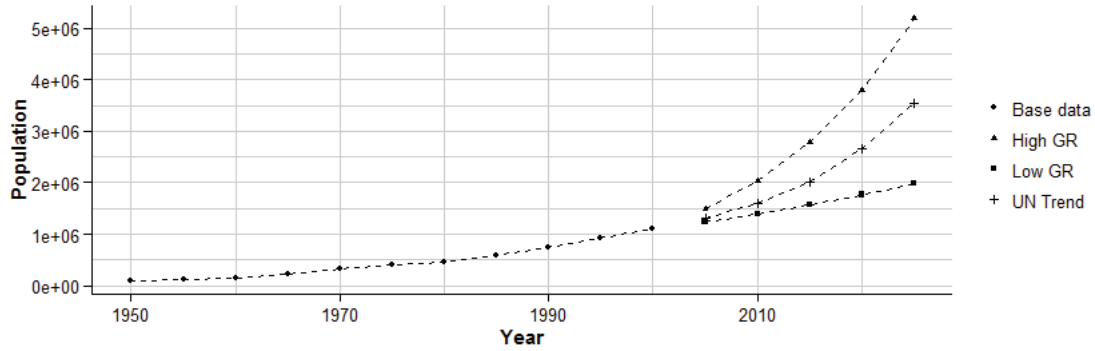


Figure 6.2: Population estimates and projections for the city of Kampala

Using the population projections reported in table 6.1 as well as the proper density, the demand for built up land can be easily estimated: it suffices to divided the population growth by the appropriate density. If the selected land use policy implies the destruction of existing buildings, such as the evacuation of flood zones, the eliminated built up land cover must be added to the demand² (see equation 6.1); results (excluding scenario specific additional demands, such as those from demolitions) are shown in table 6.1.

$$LD = \frac{(Population\ Growth)/(Population\ Density) + Demolished\ Area}{2 \times 0.04ha} \quad (6.1)$$

6.1.2 Land Use Policy Interventions and Scenario Specification

Land cover scenarios were generated by allocating the estimated land cover demands (table 6.1); for trend and high growth scenarios, the wetlands factor weight was set to 1 instead of 2 (as in the calibrated/validated model) because of expected increased pressure on suitable land by large demands for built up land. The other change relative to the calibrated/validated model was on the cells occupied by the main drainage channels of Upper Lubigi, for which land cover will not change.

Four land use policies that modify trend conditions are proposed:

- Wetlands protection: The area occupied by wetlands is constrained from further development (i.e., the cells having their centroid in areas designated as wetlands in the wetlands map will not be subject to land cover change).

²There is an implicit assumption in this reasoning, that all of the population thus displaced will be relocated within Upper Lubigi.

Table 6.1: Demographic Inputs for Scenario Construction

Parameter	Estimate		
Population data			
Total pop. of Kampala (2000)			1096690
Kawempe division share of pop. (2002 census)			22.0%
Pop. estimate of Upper Lubigi			241272
Gross density (2010) ^{1/}			554.9
Increased density (assumed) ^{2/}			1500.0
Upper Lubigi land demand estimates 2020			
	UN Trend	High growth	Low growth
Population estimate (2020)	236587	386972	82056
Land estimate (ha) - gross density	426.4	147.9	697.4
Land estimate (ha) - incr. density	355.3		

1/ Gross density: estimated population for Kawempe (under UN trend and current share) divided by the total built up area (in ha) of the Upper Lubigi catchment, 2010. Density estimates do not include as built up the area corresponding to roads. 2/ Increased density (assumed): simulated density seeking to represent urban development through redevelopment of high rise buildings in selected locations.

Source: Population data from UN Department of Economic and Social Affairs, <http://esa.un.org/unpd/wup/CD-ROM/Urban-Agglomerations.htm>; built up area derived from land cover map 2010; share of Kawempe population share based on the 2002 census data.

- Flooding area evacuation: The area currently flooded (over 5 cm depth)³ by the 1 in 10 years recurrence period storm is cleared of built up land cover (cells having their centroid in the flood areas defined with the 2010 land cover have their built up percentages and demand for built up set to 0; the built up area occupied by them in 2010 is added to the exogenous demand⁴).
- Greening cap: For built up land cover growth, the maximum accepted built up density is set at 65% (instead of the 85% used in the model calibration process); further, maximum bare soil road for these fully built up cells was set at 10%. The 2010 population density is assumed to increase by 20% (e.g., by building of a second floor in existing or new housing units). For cells with initial (2010) built up land cover exceeding 65%, no change in land cover is permitted⁵.
- Wetlands invasion: The trend constraining development on wetlands, detected in the cal-

³Damage on structures caused by a flood depth of 5 cm should be small, unless flood speed is substantive (over 2 m/s). Flood modeling does not support such condition in the case study. However, considerable smoothing has been introduced by the relatively gross size of the DEM as well as other simplifications. To partially compensate for this smoothing, a low threshold was deliberately chosen to define the flooded area.

⁴The flood model was run for the 2010 land cover map; the resulting flooded area was overlaid with the built up land cover of 2010. The built up land cover flooded was estimated as 947.6.

⁵Operationally, the process of specifying the greening cap scenario proceeded thus: the potential land cover change of the trend specification was used as a base. The potential land cover percentages were set to the 2010 value if the corresponding built up percentage of 2010 exceeded 65%. For all remaining cells with potential built up percentage over 65%, the built up percentage was changed to 65%, and (for these cells) if the bare soil percentage exceeded 10%, it was set to 10%. In order for the percentages to sum 100% (when including the water land cover), difference between the original built up/bare soil potential percentages and the capped ones, was summed to the vegetation land cover (this being the "greening" process).

ibration and validation process, is assumed to be no longer in force. The weight of the wetlands factor is set to 0.

An additional condition has also been tested: a *wetlands* invasion scenario. The logic behind this possible future is as follows: if suitable land for urban development becomes scarce, the pressure to occupy these wetlands - despite their natural propensity to flood - would likely increase. It is, therefore, to be expected under these conditions that the restriction introduced in the model for the wetland areas will no longer be operating. Thus, the scenario is defined by setting the weight of the wetland factor in the CA model to 0. Land demand will maintain trend growth and current population densities.

Table 6.2 presents the combinations of land use demand and spatial policies conforming the assessed scenarios. For each one, land cover was predicted and hydrological outcomes (total flooded area, peak discharge, flood volume and maximum depth of flood) were calculated using the open-LISEM model.

Table 6.2: Scenario specification: land demand and land use policy combinations

	Demographic Pressure		
	Trend GR	High GR	Low GR
2010 Land Cover map	S00	S07	S09
Trend Development	S01		
Flooded Areas Evacuation	S02		S10
Greening Cap	S03		
Fully Protected Wetlands	S04		
Wetlands Invasion	S05	S08	

6.2 FLOODING IMPACTS OF SIMULATED SCENARIOS

Table 6.3 presents the hydrological outcomes caused by changes in overall growth by different rates of population growth. The total built up land cover is a function of the total demand for built up land cover. High growth leads to a doubling of built up land cover, relative to 2010; the trend growth results in a similar, although somewhat smaller, level of urbanization. Low growth of population, on the other hand, only introduces a modest increase in average built up land cover.

As should be expected, relative to the 2010 base scenario, the largest impacts corresponded to the high growth scenario and the smallest, to the low growth. Interestingly, though, the increase in the peak discharge is not large: a difference of only $3.0m^3/s$, despite an almost doubling of the built up area. The total flooded area does not increase too much either; it is less than 10% for the low growth scenario and around 25% for the high growth scenario. Other impacts do grow substantially. A key impact, the total built up area affected by flooding, does expand faster than the total built up area, for example, nearly tripling for the high growth scenario.

Figure 6.4 illustrates the relationship between increases in impervious land cover and hydrological outcomes. Four groups of scenarios were generated, two corresponding to low growth population rate (S09, trend, and S10, flood area evacuation), high growth population rate (S07, trend, and S08, wetland invasion), greening (S03 with trend growth, increased density) and trend population growth (S01, trend, S02, flood area evacuation, S04, wetlands protection, and S05, wetlands

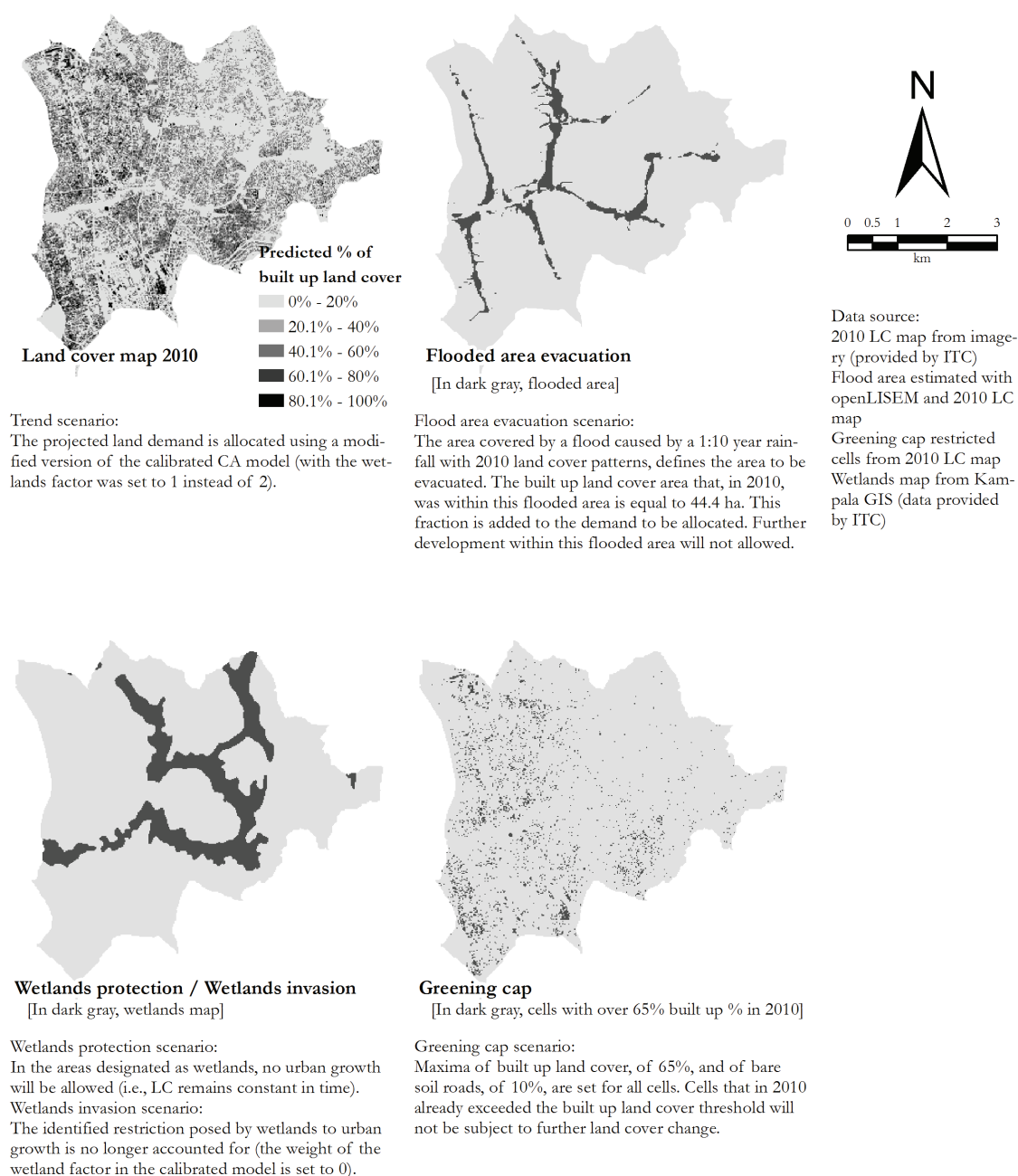


Figure 6.3: Spatial extent of spatial scenarios

invasion).

Trend growth scenarios can be used to discuss how different policy actions or dynamics drive the hydrological system. Thus, the greening scenario (S03) has a greater average percentage of vegetation and lower percentage of (impervious) built up land cover; it also has the lowest flood area, volume, peak discharge and maximum flood depth of all trend growth scenarios. Differences in peak discharge for trend growth scenarios are relatively small - all of them vary by less than 20%, relative to the 2010 base line - but other measures show larger variations. As perhaps should be

Table 6.3: Hydrological Outcomes for Simulated Trend Scenarios, 2020

	S00	S01	S07	S09
	2010 map	Trend growth	High growth	Low growth
Total built up area (<i>ha</i>)	650.09	1099.73	1291.58	801.04
Affected built up area (<i>ha</i>)	37.90	76.64	101.72	50.81
	5.83%	6.97%	7.88%	6.34%
Peak discharge (m^3/s)	56.00	57.78	59.01	56.66
Flood Volume (m^3)	542.5×10^3	772.9×10^3	885.6×10^3	596.7×10^3
Max flood level (<i>m</i>)	1.56	2.15	2.35	1.64
Total flood area (over 5 cm depth, <i>ha</i>)	179.12	209.96	223.08	187.20

Flooding results simulated with openLISEM; rainfall corresponds to approximately 1:10 return period.

expected, protecting the wetlands from further urbanization (S04) leads to smaller flood volume and (slightly) peak discharge. The trend scenario (S01), the potential occupation of wetlands (S05) and the evacuation of the flooded area (S02) cause very similar levels of both peak discharge and flood volume.

In terms of total flood area and maximum flood depth, S01, S02 and S04 all show similar levels, and S05 (wetlands invasion) seems to present lower impacts. Flood areas and maximum depth levels are likely associated to bottlenecks and obstructions in the drainage system; by allowing the urbanization in the areas nearest to the channels, S05 could lead to faster water flows. This is not a positive outcome; while apparently reducing impacts, faster water flows are more dangerous and the mitigation obtained from less flooded area is likely to be lost to higher damages caused by the flood.

When considering low growth (S09 and S10) and high growth (S07 and S08) scenarios, flooded area, maximum flood depth and peak discharge all result in equal measures. The flood volume of the low growth trend (S09) is lower than for S10, evacuation of the 2010 flood area; inversely, the high growth trend conditions, S07, cause a larger flood volume than when wetlands are freed for development - perhaps because the latter areas are in low lying relief, closer to the drainage system, and the generated runoff of this new built up land can be evacuated faster. Still, the difference between S07 and S08 is small.

Figure 6.5 seeks to explain how certain hydrological outcomes, such as the total flooded area and peak discharge, impact the land cover pattern, in terms of the flooded built up area (in relative and absolute terms). Evidently, the affected area is lowest for low population growth and highest for high population growth. It is very interesting to note that, while the evacuation of the 2010 flooded area does reduce significantly the impact (S10 has around one fifth of the flooded area of 2010), this is not true of S2 (same evacuation of flooded area but under trend conditions): the total built up area affected is equal to that of 2010 (of course, in relative terms, there is an improvement, since the total built up area has increased substantially).

Another interesting trend: despite increases of the total flooded area, the level of affected built up percentage for low growth trend (S09) and trend growth (S03, S05, S01, S04) is essentially constant, at a ratio to 2010 that is slightly higher than 1. In absolute terms, the affected area doubles for three trend growth scenarios (S01, S04, S05). The greening scenario (S03) results in a nearly 50% increase in the total affected area, smaller than the other trend scenarios because the larger density

implicit in the scenario creates overall less built up land cover. In general, the impacts do not seem associated with flood area and peak discharge, perhaps with the exception of peak discharge and average built up percentage - which shows a direct relationship (as peak discharge grows, so does the percentage of built up percentage).

6.3 DISCUSSION OF SCENARIO SIMULATIONS

It is very clear that land cover patterns do impact the level of flooding within Upper Lubigi. Less clear is how large an impact on reducing overall levels of affected (by flooding) built up land cover. Rather than the pattern, the most important factor seems to be the overall level of development within the study area.

A policy of flood evacuation may be very effective in mitigating flood problems in Upper Lubigi if population growth is not too large. But under trend conditions, it merely stabilizes the problem. A key question that emerges is, then, if alternatives to this policy should be considered. In particular, guiding future urban development to other areas of Kampala and away from Upper Lubigi may result in similar mitigation effects of flooding affectation. This is not to say that the location of new development is irrelevant; for example, changing the built form through an intensification of land use (e.g. higher population density, as in S03) certainly contributes to mitigate flood problems (in S03, by reducing the total amount of required land). Scenarios were chosen to explore bounding conditions, the limits of land policy; but combinations of the measures discussed, and of other complementary actions (in terms of infrastructure investment, both in households and in the drainage system), can be designed to manage runoff in a more sustainable and efficient manner.

A second conclusion of interest refers to the protection of wetlands. Under high pressure (very high population growth), similar patterns of land cover and of affected built up land cover are the result. The historic protection of wetlands may result in some mitigation of flooding (less total flooded area, peak discharge, although even under trend conditions, the affected built up area of the trend, S01, is essentially equal to wetlands protection, S04). On the positive side, purely in terms of flooding impacts, further occupation of wetlands by built up land cover does not generate large flooding impacts (the pairs of scenarios S07 and S08, high growth, and S01 and S04, trend growth, both show similar levels of affectation). In the same line, protecting the wetland areas does not seem to reduce flooding impacts substantially.

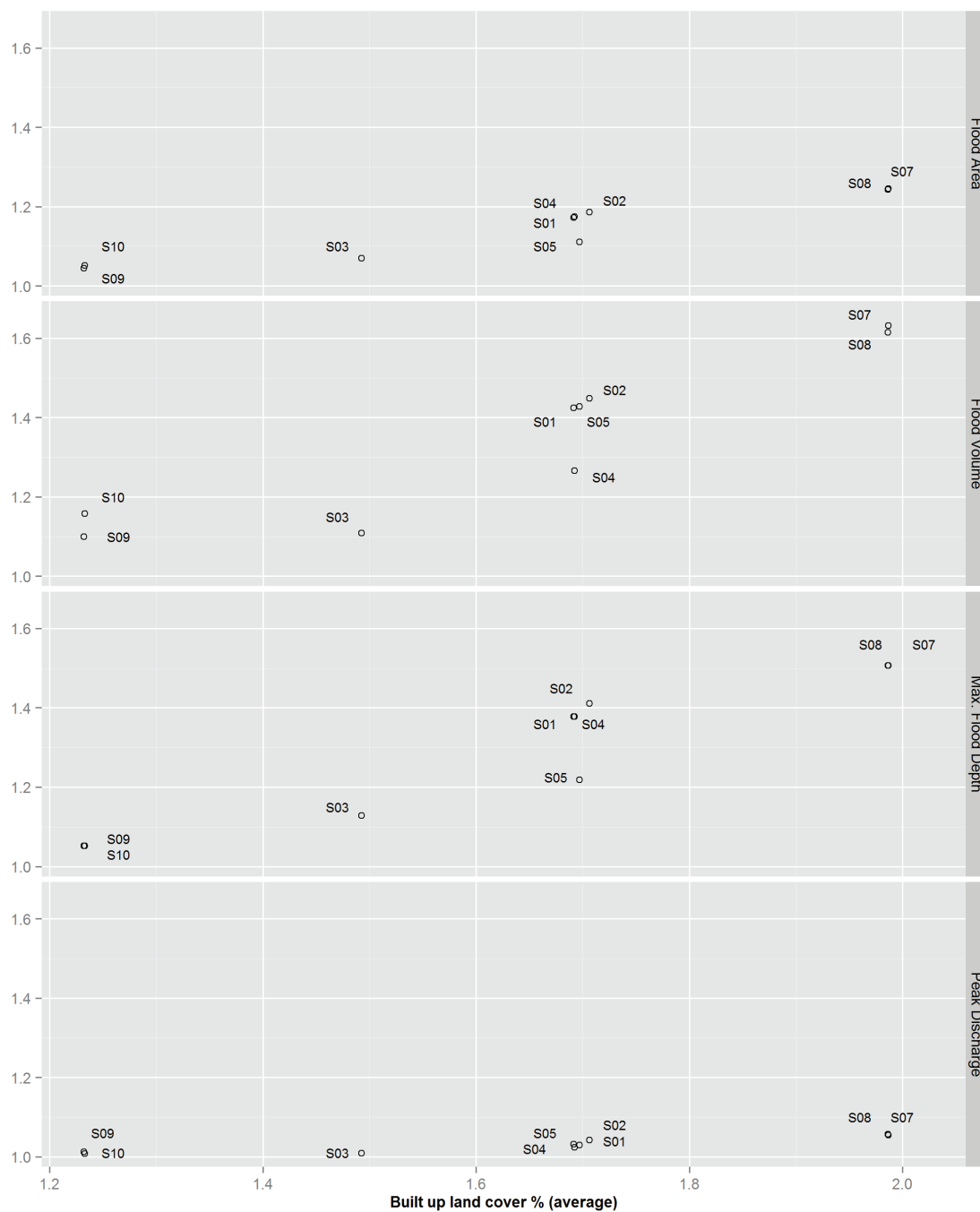


Figure 6.4: Ratio of scenario value to 2010 conditions. Hydrological outcomes vs. average percentage of built up land cover

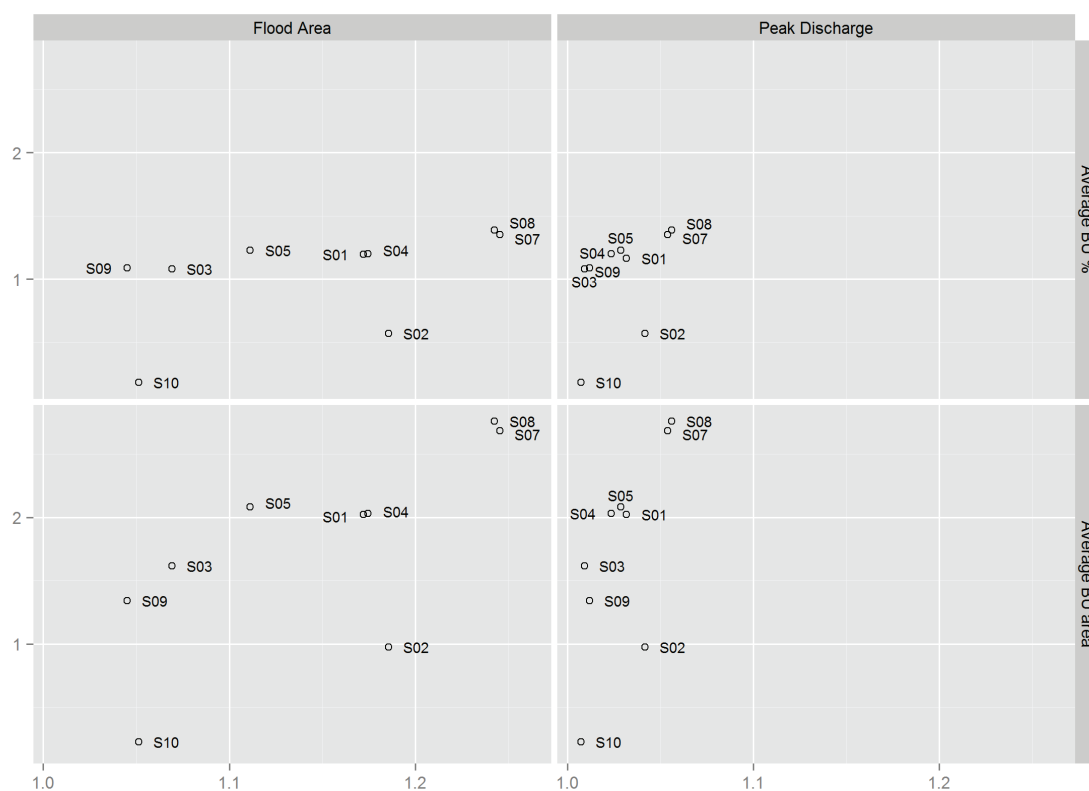


Figure 6.5: Ratio of scenario value to 2010 conditions. Flooded area vs. hydrological outcomes

Chapter 7

Concluding Remarks

This research exercise is an application of spatially explicit modeling techniques for designing solutions to specific spatial problems. A range of scenarios and the calibrated CA model were used to simulate possible future land cover patterns for Upper Lubigi. These scenarios included variations on the demand for built up land (in turn, derived from different population projections) and on land use policies (the evacuation of areas currently flooded, the protection from future development of identified wetlands, an increase in population density and a cap on the amount of built up land cover in each cell, and the elimination of constraints on wetland development).

For each scenario, hydrological outcomes were predicted under constant soil properties and for a 1:10 years (approximately) return period storm. In particular, maps of flood depth were generated using a calibrated openLISEM model of Upper Lubigi. These were overlayed on top of the simulated land cover patterns to estimate the built up area flooded under the scenario conditions. This process is a soft coupling of the calibrated CA model with the openLISEM flood model; the outputs of the CA model were introduced as inputs into the openLISEM flood model.

The CA model was developed by identifying the subset of factors which, combined by means of a weighed summation, best replicated the 2004-2010 growth in built up land cover for Upper Lubigi. The calibration process included (1) the exploration of the relationship between each factor and the increase in built up land cover, (2) the selection of the best predictive factors, (3) the selection of weights (relative importance of each factor), and (4) the introduction of additional restrictions (the constraint of development on areas occupied by institutional land uses). The resulting model was validated by predicting the 2010 built up land cover of Nalukolongo, from the 2004 land cover map of this subcatchment, and comparing this prediction to the 2010 land cover map. The model was found to perform in a satisfactory manner, perhaps even better than for the calibration data.

Land cover data models (maps) of Upper Lubigi were assessed and improved. The existing feature data set of buildings areas of 2004 was visually inspected; missing polygons were copied from the 2010 data set, digitized or adopted from an unsupervised classification of the imagery. A topological revision of the feature data set of building area reduced problems of overlapping polygons in both 2004 and 2010. The land cover maps were extended to the entire Lubigi subcatchment, as well as to Nalukolongo, with similar levels of accuracy to that of previous classifications (in particular, the work of Fura, 2013). This expansion of the extent also included the area of Upper Lubigi located beyond the limits of the KCCA (this last extension was only used in the simulation process, not in the calibration phase).

7.1 LAND COVER MODELING AND SIMULATION OF SCENARIOS

The total amount of development within Upper Lubigi was identified as the key variable determining the level of impact. Flooding impacts were found to be much more sensitive to differences in

population growth rates than on conditions determining the arrangement of land cover classes in the landscape. Even under greening conditions, no major differences were introduced (the reduction in impact of such a scenario is associated with a higher density, i.e., lower land demand rather than with any spatial differential on the land cover pattern). However, structural solutions were not simulated, neither on the drainage system and public infrastructure nor on housing units. Existing evidence strongly suggests that such measures produce major reductions in flooding impacts (Mohnda, 2013; Sliuzas, Flacke, & Jetten, 2013).

The evacuation of flooded areas, a typical policy response, merely ameliorates the flooding problems in the long run: under trend conditions of spatial distribution and population growth, flooding impacts were equal to the base year ten years into the simulation. Only with very low population growth was a substantial reduction of impact achieved. This suggests efforts should concentrate on reducing the amount of runoff (mitigating the flood hazard) rather than rearranging existing development (adapting to hazard conditions).

Determinants of land cover were found to be in line with both previous evidence on Kampala (Vermeiren et al., 2012; Abebe, 2013; Fura, 2013) and theoretical expectations: built up land cover increases were simulated based on a neighborhood effect (local context), non-vegetation land cover of the base year (inertia, i.e., resistance to change) and estimated travel time to the CBD (accessibility). Interestingly, wetlands were found to pose a relatively strong disincentive to development.

No feedback effects between flooding and urban development were detected. The differences in time and spatial scales of the phenomena involved, local conditions and lack of sufficiently detailed data (e.g., historical data on rainfall, higher temporal resolution of land cover data models) may explain this absence.

7.2 METHODOLOGICAL ISSUES

The model was generally successful in replicating past dynamics, both in the calibration and the validation areas (perhaps even more so in Nalukolongo, the validation area, than in Upper Lubigi; this could be explained by differences in the land cover patterns of these areas, although such point was not a major concern of this study). Further, the simulations of potential futures were an improvement on previous results, based solely on statistical analysis. While over-clustering remains in the simulated results, this problem has been much reduced relative to Logit model-based projections of land cover patterns. The introduction of randomness into the simulation process was more successful than in previous efforts (Vermeiren et al., 2012; Fura, 2013).

Perhaps unsurprisingly, the amount of information (factors explaining land cover) had a greater impact on calibration than the relative importance (weight) of each factor. Heuristic analysis and conformity to theory remain key considerations in selecting proper factors. As a general principle, the least amount of factors should be introduced into a specific model, to guard against overfitting.

Statistical assessment of land cover patterns was useful in characterizing the influence of individual factors (i.e., the differences between predictions based on single factor models) but much less so in comparing different combinations of factors. Statistical assessment may be more properly used to understand relationships than to optimize the model calibration.

Randomness was introduced through the simulation of future land cover. A major lesson of the analysis of land cover patterns was that the *location* of built up land cover growth can be attributed to substantive determinants; but the *amount* of growth in each cells seems to be essentially random. It is difficult to ascertain whether this is due to imperfections of the dependent variable (built up

land cover area vs. more sophisticated measures, such as density of population, households or housing units), uncertainty in the original land cover data models, a inherent characteristic of the city or a combination of these (and other) circumstances.

The formalization of space as a collection of discrete features (in this application, square cells) may be a very useful characteristic of the model. The success of the modeling efforts proves it can successfully represent land cover patterns using a regular arrangement. But it also has the potential to improve analysis in areas where more information exists; in particular, in land systems where property relationships are strong determinants of urban growth, the use of cadastral boundaries will likely be the proper spatial unit of analysis. The chosen formalization of space could accommodate such constraints (such an adaption would require, mostly, new algorithms to update the neighborhood effect because in the current version, it relies of a square moving window and its accord with the square feature data set underlying the model).

The model and simulations developed have made use of deliberately few sources of information, generally available in even in data scarce environments. This information has mostly been: general demographic data (population trends and low resolution spatial disaggregation of population distribution), high spatial resolution remote sensing imagery (but with low spectral resolution: only the three true bands), and general maps of roads and wetlands (the latter at a very low spatial detail); in addition, a digital elevation model, soils information and political divisions of and within the city. Such information is characterized by high levels of uncertainty, which propagates through modeling and simulation, thus the need for a scenario approach. But it does have the advantage of being useful for rapid assessments and for the analysis of cities with weak institutional contexts or where few spatial information exists.

Because no feedback was detected, a hard coupling of the CA model and the flood model was not attempted. However, while the details of implementation may be time consuming, such a coupling is relatively straightforward: the CA model was developed in ArcGIS 10.1 ModelBuilder Model. It can be exported as a Python script and run from this alternative platform. openLISEM, the flood model, is based on the relatively low level PC Raster platform. But PC Raster functions may also be called from the higher level Python. The key concerns in this process is the transformation of the tables, resulting from the CA model, into PC Raster format as well as the reverse, the exporting of the PC Raster flood maps into the feature data set of ArcGIS (in this project, this step was manually performed; it is trivial).

The restriction on land cover change introduced by institutional land uses contributed to improve the predictions of the model. Unlike the traditional constraints approach, which modifies the suitability value of the cells within the constrained space, the selected method relies on changing the potential simulated land cover. This approach is similar to that of the METRONAMICA software¹. This specification provides flexibility to the model: areas remaining constant may be land covers (tarmac), land uses (institutional) or even policy choices (e.g. protected areas, as in the wetland protection scenarios). With more detailed land use models of Kampala, restricted areas could be refined and improved but also more dynamics could be introduced (e.g. different development types could be treated differently, expanding the model capabilities). Additionally, suitability constraints (setting the suitability of certain areas to a 0 value) may also be implemented jointly with the selected method.

¹METRONAMICA divides the possible land use categories into features, that do not change in the simulation, and functions and vacant (the former, the land uses being simulated; the latter, changing in response to the functions). In the CA model design, built up land cover would be akin to the function; tarmac land cover and institutional land uses, similar to the features, and other land covers, analogous to the vacant (their change is determined by the dynamics of the built up land cover, whether it be increases - on road bare soil - or decreases - off road bare soil and vegetation).

7.3 OPPORTUNITIES FOR FURTHER RESEARCH

The results of this project clearly support the success of the general approach of spatially explicit scenario development in this case study. However, a number of limitations have been identified. They constitute interesting avenues to extend and improve the general method as well as the results for Kampala.

- **Emerging properties:** The implementation of an automated feedback loop between flooding and urban development would allow for a potential emerging property within the model. This seems to be unnecessary for Kampala but may be of general interest.
- The application of the modeling framework to other case studies (other cities) would allow for a better exploration of the range of conditions under which it is applicable, as well as the changes required and the sensitivity to local context of the general approach (this is relevant for the previously said feedback but also for many other constraints and opportunities). In general, the framework is promising for local flooding (as opposed to large river systems inundation) and in situations of urban growth. Open questions include the influence on the model of: changes in climatic conditions over the long run, the interaction of local and higher scale flooding, low rates of urban expansion, rapid changes in accessibility, stronger policy and legal constraints.
- The calibration and simulation with the model of areas richer in information may contribute both to more sophisticated applications and to test the sensitivity of the model to more variables. In particular, the use of historic rainfall data (and other hydrological information), more detailed population data (spatially and in time), remote sensing data (especially an infrared band to better characterize bare soil and vegetation).
- For the specific case of Kampala, the changes in built up land cover have been well studied. However, more information and analysis would be convenient to better understand other land cover transitions and boundaries (e.g., between vegetation and bare soil, between bare soil and tarmac, between types of vegetation). Eventually, built up land cover classes may also be distinguished - although this last may require substantial reformulation of the CA model.
- The assessment of accord between land cover maps and predictions, in the calibration and validation stages, is also a fruitful area for further development. Two possible ways forward include the application of fuzzy methods and the evaluation of prototypical sites within the study area, predefined before the calibration/validation is attempted.
- The coupling of the CA model and the flood model should be generalized to include other effects - on a case by case basis - such as: the operation of infrastructure systems (e.g., drainage channels, road congestion) and the impact of infrastructure investment (on operation but also on attractiveness for development).
- Specifically for Kampala, the range of scenarios simulated may be profitably expanded beyond general land cover differentials. Infrastructure and local infiltration was mentioned as a promising area; redevelopment at higher density, with its associated road (tarmac cover) changes may also be of interest. Additionally, other policy goals should be considered, as well as impact measures to assess them. Efforts to identify tradeoffs between goals are particularly important.

References

- Abebe, G. A. (2013). *Quantifying urban growth pattern in developing countries using remote sensing and spatial metrics: A case study in Kampala, Uganda*. Unpublished master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.
- Alonso, W. (1964). *Location and land use. Toward a general theory of land rent*. Cambridge, MA: Harvard University Press.
- Batty, M., & Torrens, P. M. (2005). Modelling and prediction in a complex world. *Futures*, 37(7), 745–766. doi: 10.1016/j.futures.2004.11.003
- Benenson, I., & Torrens, P. M. (2004). *Geosimulation: Automata-based modeling of urban phenomena*. Chichester, UK: Wiley & Sons.
- Brueckner, J. K. (1987). The structure of urban equilibria: A unified treatment of the Muth-Mills model. In E. S. Mills (Ed.), *Handbook of Regional and Urban Economics. Volume II* (pp. 821–845). Amsterdam: North Holland.
- Cheng, J., & Masser, I. (2003). Modelling urban growth patterns: a multiscale perspective. *Environment and Planning A*, 35(4), 679–704. doi: 10.1068/a35118
- Ciavola, S. J., Jantz, C. A., Reilly, J., & Moglen, G. E. (2012). Forecast Changes in Runoff Quality and Quantity from Urbanization in the DelMarVa Peninsula. *Journal of Hydrologic Engineering*. (Accepted for publication) doi: 10.1061/(ASCE)HE.1943-5584.0000773
- Clark, W. C. (1980). Witches, floods, and wonder drugs: historical perspectives on risk management. In R. C. Schwing & W. A. Albers (Eds.), *Societal risk assessment: How safe is safe enough* (pp. 287–313). New York: Plenum Press.
- Clarke, K. C., Gazulis, N., Dietzel, C. K., & Goldstein, N. C. (2007). A decade of SLEUTHing: Lessons learned from applications of a cellular automaton land use change model. In P. Fisher (Ed.), *Classics from IJGIS. Twenty Years of the International Journal of Geographical Information Systems and Science* (pp. 413–425). Boca Raton, FL: Taylor and Francis, CRC.
- Clarke, K. C., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay Area. *Environment and Planning B*, 24(2), 247–261. doi: 10.1068/b240247
- Cronshey, R., McCuen, R. H., Miller, N., Rawls, W., Robbins, S., & Woodward, D. (1986). *Urban Hydrology for Small Watersheds* (Tech. Rep. No. TR-55). Washington DC: US Department of Agriculture, Soil Conservation Service, Engineering Division.
- De Roo, A., Wesseling, C., & Ritsema, C. (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. doi: 10.1002/(SICI)1099-1085(199608)10:8<1107::AID-HYP415>3.0.CO;2-4

- Echols, S. P., & Nassar, H. F. (2006). Canals and lakes of Cairo: influence of traditional water system on the development of urban form. *Urban Design International*, 11(3), 203–212. doi: 10.1057/palgrave.udi.9000176
- Eigenbrod, F., Bell, V., Davies, H., Heinemeyer, A., Armsworth, P., & Gaston, K. (2011). The impact of projected increases in urbanization on ecosystem services. *Proceedings of the Royal Society B: Biological Sciences*, 278(1722), 3201–3208. doi: 10.1098/rspb.2010.2754
- Ermentrout, G. B., & Edelstein-Keshet, L. (1993). Cellular automata approaches to biological modeling. *Journal of theoretical Biology*, 160(1), 97–133. doi: 10.1006/jtbi.1993.1007
- Fang, S., Gertner, G. Z., Sun, Z., & Anderson, A. A. (2005). The impact of interactions in spatial simulation of the dynamics of urban sprawl. *Landscape and Urban Planning*, 73(4), 294–306. doi: 10.1016/j.landurbplan.2004.08.006
- Fura, G. D. (2013). *Analysing and modelling urban land cover change for run-off modelling in Kampala, Uganda*. Unpublished master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.
- Hu, S., & Bing, H. (2011). The effect of urban flood control on evolution of the urban morphology: Case study of Changde, Hunan (*In Chinese*)(*English abstract*). In *2011 International Conference on Multimedia Technology (ICMT)* (pp. 4182–4185). doi: 10.1109/ICMT.2011.6002871
- Huong, H. T. L., & 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. doi: 10.5194/hess-17-379-2013
- Jain, D. (2009). *Studying and Modelling Changing Urban Form*. Unpublished master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, The Netherlands.
- Jetten, V. G. (2013). *openLISEM flood hazard simulation in Kampala*. Retrieved from <http://www.youtube.com/watch?v=APiRzhV0h8c&feature=youtu.be> (Accessed: Aug. 8, 2013)
- Jha, A., Lamond, J., Bloch, R., Bhattacharya, N., Lopez, A., Papachristodoulou, N., ... Barker, R. (2011). *Five Feet High and Rising. Cities and Flooding in the 21st Century* (Policy Research Working Paper No. 56481). Washington DC: The World Bank.
- Koomen, E., & Stillwell, J. (2007). Modelling land-use change. In E. Koomen, J. Stillwell, A. Bakema, & H. Scholten (Eds.), *Modelling land-use change*. Springer Netherlands.
- LeSage, J. P., & Pace, R. K. (2009). *Introduction to Spatial Econometrics*. Boca Raton, FL: CRC Press.
- Li, X. (2011). Emergence of bottom-up models as a tool for landscape simulation and planning. *Landscape and Urban Planning*, 100(4), 393–395. doi: 10.1016/j.landurbplan.2010.11.016
- Li, X., & 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. doi: 10.1080/136588100240886
- Lwasa, S. (2010). Adapting urban areas in Africa to climate change: the case of Kampala. *Current Opinion in Environmental Sustainability*, 2(3), 166–171. doi: 10.1016/j.cosust.2010.06.009

- Maithani, S. (2010). Application of Cellular Automata and GIS Techniques in Urban Growth Modelling: A New Perspective. *Institute of Town Planners, India Journal*, 7(1), 36–49.
- McHarg, I. L. (1969). *Design with Nature*. New York: American Museum of Natural History.
- McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science*, 32(3), 335–348. doi: 10.1111/j.1467-9787.1992.tb00190.x
- Mejía, A. I., & Moglen, G. E. (2009). Spatial patterns of urban development from optimization of flood peaks and imperviousness-based measures. *Journal of Hydrologic Engineering*, 14(4), 416–424. doi: 10.1061/(ASCE)1084-0699(2009)14:4(416)
- Mohnda, A. (2013). *Evaluating Flash Flood Risk Reduction Strategies in Built-up Environment in Kampala*. Unpublished master's thesis, Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, Enschede, the Netherlands.
- Parker, D. J. (1995). Floods in Cities: Increasing Exposure and Rising Impact Potential. *Built Environment*, 21(2-3), 114–125.
- Poelmans, L., Rompaey, A. V., Ntegeka, V., & 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. doi: 10.1002/hyp.8047
- R Core Team. (2013). R: A Language and Environment for Statistical Computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Reams, Z. (2010). *Kappa Statistics v1.1*. <http://arcscrippts.esri.com/details.asp?dbid=16795>.
- Santé, I., García, A. M., Miranda, D., & 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. doi: 10.1016/j.landurbplan.2010.03.001
- Sliuzas, R., Flacke, J., & Jetten, V. (2013). Modelling urbanization and flooding in Kampala, Uganda. In *Proceedings of the 14th N-AERUS / GISDECO conference*. Enschede, the Netherlands.
- Sliuzas, R., Lwasa, S., Jetten, V., Petersen, G., Flacke, J., & Wasige, J. (2013). Searching for Flood Risk Management Strategies in Kampala. In *Proceedings of the 14th N-AERUS / GISDECO conference*. Dublin, Ireland.
- Soares-Filho, B., Rodrigues, H., & Follador, M. (2013). A hybrid analytical-heuristic method for calibrating land-use change models. *Environmental Modelling & Software*, 43, 80–87. doi: 10.1016/j.envsoft.2013.01.010
- Soares-Filho, B. S., Coutinho Cerqueira, G., & Lopes Pennachin, C. (2002). DINAMICA—a stochastic cellular automata model designed to simulate the landscape dynamics in an Amazonian colonization frontier. *Ecological Modelling*, 154(3), 217–235. doi: 10.1016/S0304-3800(02)00059-5
- Steinitz, C., Arias, H., Bassett, S., Flaxman, M., Goode, T., Maddock, T., ... Shearer, A. (2003). *Alternative futures for changing landscapes: the upper San Pedro River Basin in Arizona and Sonora*. Covelo Island, CA and Washington DC: Island Press.

- Steinitz, C., Faris, R., Flaxman, M., Vargas-Moreno, J. C., Canfield, T., Arizpe, O., ... others (2005). A sustainable path? Deciding the future of La Paz. *Environment: Science and Policy for Sustainable Development*, 47(6), 24–38. doi: 10.3200/ENVT.47.6.24-38
- Timmermans, H. (2012). On the Simplicity of Complexity Theory in Artificial Environments. In *Complexity Theories of Cities Have Come of Age* (pp. 173–184). Berlin & Heidelberg: Springer.
- Tobler, W. R. (1979). Cellular geography. In S. Gale & G. Olsson (Eds.), *Philosophy in geography* (pp. 379–386). Dordrecht, the Netherlands: D. Reidel Publishing Company.
- Toffoli, T., & Margolus, N. (1987). *Cellular automata machines: A new environment for modelling*. Cambridge, MA: MIT press.
- van Delden, H., Escudero, J. C., Uljee, I., & Engelen, G. (2005). METRONAMICA: A dynamic spatial land use model applied to Vitoria-Gasteiz. In *Virtual Seminar of the MILES Project*.
- van Vliet, J., Hagen-Zanker, A., Hurkens, J., & van Delden, H. (2013). A fuzzy set approach to assess the predictive accuracy of land use simulations. *Ecological Modelling*, 261-262, 32–42. doi: 10.1016/j.ecolmodel.2013.03.019
- van Vliet, J., Hurkens, J., White, R., & van Delden, H. (2012). An activity-based cellular automaton model to simulate land-use dynamics. *Environment and Planning B*, 39(2), 198–212. doi: 10.1068/b36015
- Vargas-Moreno, J. C., & Flaxman, M. (2012). Using participatory scenario simulation to plan for conservation under climate change in the greater everglades landscape. In H. A. Karl, L. Scarlett, J. C. Vargas-Moreno, & M. Flaxman (Eds.), *Restoring lands-coordinating science, politics and action* (pp. 27–56). Dordrecht, the Netherlands: Springer.
- Vermeiren, K., van Rompaey, A., Loopmans, M., Serwajja, E., & Mukwaya, P. (2012). Urban growth of Kampala, Uganda: Pattern analysis and scenario development. *Landscape and Urban Planning*, 106(2), 199–206. doi: 10.1016/j.landurbplan.2012.03.006
- White, R. (1998). Cities and cellular automata. *Discrete dynamics in Nature and Society*, 2(2), 111–125. doi: 10.1155/S1026022698000090
- White, R., & Engelen, G. (1993). Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns. *Environment and Planning A*, 25(8), 1175–1175. doi: 10.1068/a251175
- White, R., Uljee, I., & Engelen, G. (2012). Integrated modelling of population, employment and land-use change with a multiple activity-based variable grid cellular automaton. *International Journal of Geographical Information Science*, 26(7), 1251–1280. doi: 10.1080/13658816.2011.635146
- Wu, F. (1998). SimLand: a prototype to simulate land conversion through the integrated GIS and CA with AHP-derived transition rules. *International Journal of Geographical Information Science*, 12(1), 63–82. doi: 10.1080/136588198242012
- Wu, F., & Webster, C. J. (2000). Simulating artificial cities in a GIS environment: urban growth under alternative regulation regimes. *International Journal of Geographical Information Science*, 14(7), 625–648. doi: 10.1080/136588100424945

- Yang, G., Bowling, L. C., Cherkauer, K. A., & Pijanowski, B. C. (2011). The impact of urban development on hydrologic regime from catchment to basin scales. *Landscape and Urban Planning*, 103(2), 237–247. doi: 10.1016/j.landurbplan.2011.08.003
- Yeh, A. G.-O., & 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. doi: 10.1068/b2740
- Yeh, A. G.-O., & Li, X. (2002). A cellular automata model to simulate development density for urban planning. *Environment and Planning B*, 29(3), 431-450. doi: 10.1068/b1288
- Zare, S. O., Saghafian, B., & Shamsai, A. (2012). Multi-objective optimization for combined quality–quantity urban runoff control. *Hydrology and Earth System Sciences*, 16(12), 4531–4542. doi: 10.5194/hess-16-4531-2012

Appendix A

Land Cover Data Model Development

This appendix summarizes the development of land cover data models, from a set of input maps previously developed by Fura (2013) as well as the raw information available to him. The objective is to create a description of land cover for the years 2004 and 2010 that may be used to model urban growth and flooding in the catchments of Lubigi and Nakulolongo.

A.1 DATA INPUTS

Data inputs used for the development of final land cover models include three sets of information:

1. Existing land cover maps
 - (a) Land cover and road maps developed by Fura (2013) for the Lubigi catchment
 - (b) Land use maps of the KCCA for 2004 and 2010
2. Unprocessed imagery: mosaic of 2004 with a spatial resolution of 0.623 m and aerial image of 2010 with a spatial resolution of 0.5. The spectral resolution of both images encompasses only three bands of the visible spectrum.
3. Building footprints maps (vectorized data) of the KCCA for 2010 and 2004

Figure A.1 shows the KCCA, the study area and the extent of the various data inputs.

A.2 QUALITY ASSESSMENT OF EXISTING DATA MODELS

Data needs for the development of a final land cover change map are to be assessed based on two criteria: accuracy and extent. Based on these, a strategy for the improvement of these data models was proposed and executed.

A.2.1 Extent

Figure A.1 shows the extent of the main data inputs. The basic data for building footprints, imagery and ancillary information is available for the entire KCCA and, therefore, does not represent a problem in terms of its eventual use in creating a final land cover change model.

These initial data sets were substantially improved by Fura (2013). However, as can be seen, his study area only includes the northern portion of the total area needs for this project. As such, the data generated by Fura cannot be directly used. Rather, the insights developed in his project will

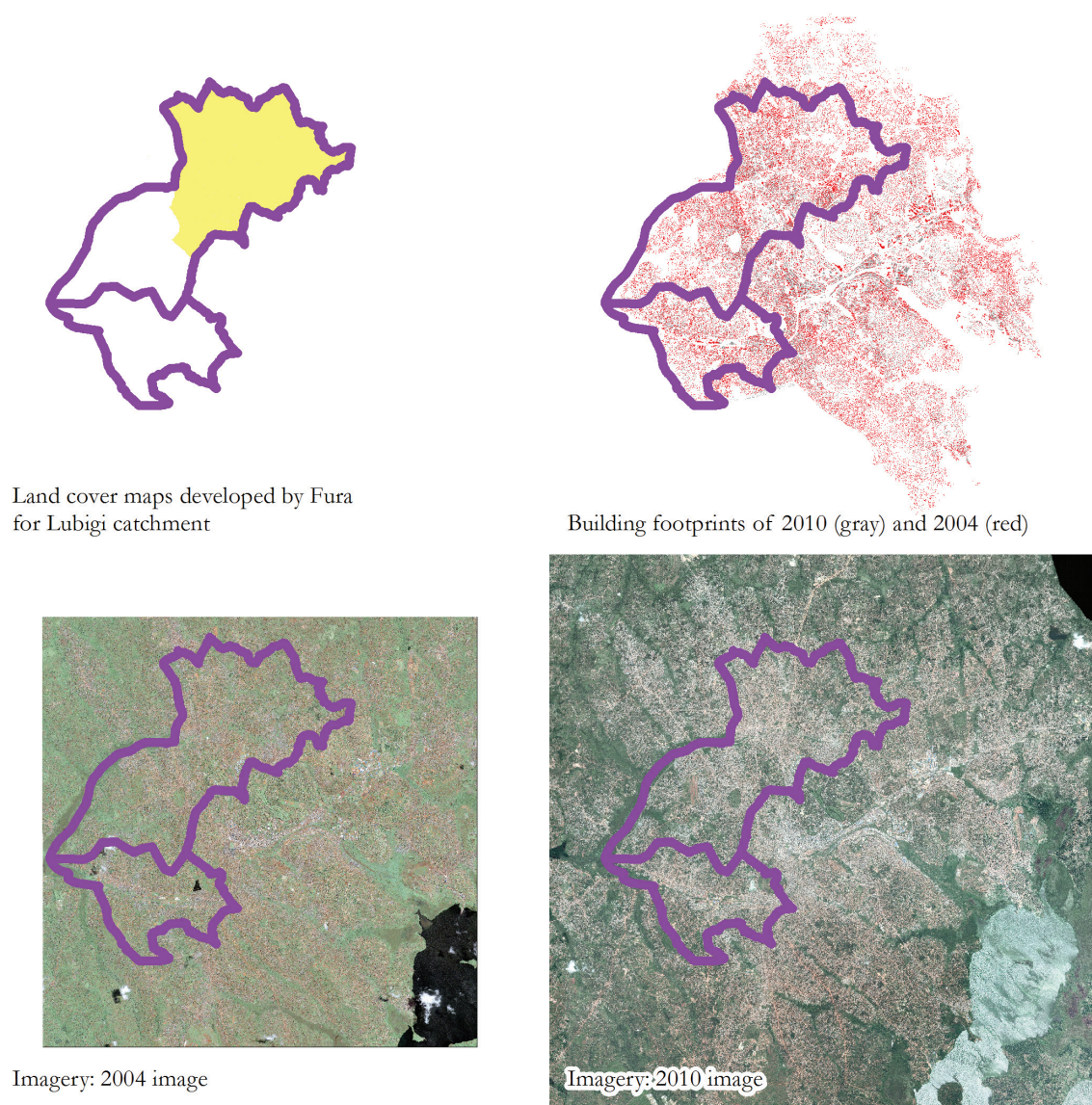


Figure A.1: Extent of input data models for land cover development

inform the classification to be applied. Particularly, Fura (2013) detected the existence of a bias in the 2004 building footprints identification, resulting from a confusion between the brown roofs of certain buildings and the color of bare soil. (During the assessment of building footprints during this project, the opposite bias was detected for 2010: the overlapping of polygons representing built up structures; however, since all information was rasterized, this error was judged to introduce no bias in the modeling process of the land cover.)

A.2.2 Accuracy

An accuracy assessment of the original building footprints (vector data) and the classification by Fura (2013) in terms of built up/non built categories, was conducted, to better understand the state of existing land cover maps.

Given land cover maps of two categories (built up/non built) of two periods, four possible cases exist:

1. Area classified as built up in both years: it is very likely built up, although it might be confused with bare soil
2. Area classified as built up in 2004 and not in 2010: two options, an error (it is really non built in 2004 or built up in 2010) or a demolition (relatively unlikely although logically possible)
3. Area classified as built up in 2010 and not built up in 2004: two options, urban growth (new construction) or an error (non built up in 2010 or built up in 2004)
4. Area classified as not built up in both years: two options, if it is vegetation, it almost certainly is not built up; on the other hand, bare soil may be confused with built up in both years

A confusion matrix was estimated to quantify the accuracy level of this classification. It is based on a stratified sample of point locations. The stratification was based on the four mentioned cases, as well as the location of each point within either of the catchments of the study area. Thus, 100 points were sampled in random locations for zones corresponding to each category (e.g. areas that were built up in 2004 but not built up in 2010; these maps were generated by overlaying the two building footprints vectorized data). Half the points were sampled in Nakulolongo and half in Lubigi.

For each point, land cover status (built up or non built) was determined by visual inspection of the imagery. This evaluation was done at a scale of approximately 1:1200, separately for each year. For border line cases, the data model (building footprints map) was generally assumed to be correct. For example, if a point was located in the shadow cast by a building and the land cover model identified this area as not built, the point was considered in the not built category.

The points located within the area for which Fura produced a land cover map were extracted and analyzed separately. These are reported in table A.2.

Three conclusions can be immediately drawn from considering these tables.

Firstly, as reported in Fura (2013), there is a severe problem of underestimation with the 2004 building footprints. Of 100 sampled points, only 13 were correctly classified as being only built up in 2004. This points to the rarity of demolitions, in general, but more importantly to the fact that building footprints of the year 2004 were not completely extracted in 2004.

A problem of underestimation was also detected for 2010. Fully 46% of locations were classed as being only built up for 2010 in the map and as built up for both periods in the sample. But this statistic is somewhat misleading, as it seems to be more a consequence of the sampling rather than a substantive discrepancy: because of the stratification strategy, there are 100 points in locations classed by the maps in each category, regardless of the area occupied by each category.

The category built up only in 2010 consists mostly of narrow slivers. For many cases, the 2004 building is right next to the data point, close enough to attribute the problem to a displacement between the images but not close enough to accurately classify it as built up in 2004. In terms of data needs, these small errors are inconsequential because the land cover maps will be transformed into percentages of larger (10 m sided) cells.

Secondly, the procedure outlined and executed by Fura (2013) was successful in improving the general land cover pattern. Overall accuracy increased by 27% - from 56% to 72% (these percentages seem low but it must be recalled that the sample was deliberately built to give equal weight to land

Table A.1: Confusion matrix. Accuracy assessment of land cover change in original building footprints, 2004-2010

Location is built up in period:							
Data from maps	Data from sample					Total	User's acc.
	Both	None	Only 2004	Only 2010	Unclass.		
Both	83	2		14	1	100	83.8%
None	4	84	1	11		100	84.0%
Only 2004	42	32	13	11	2	100	13.3%
Only 2010	45	11		44		100	44.9%
Total	174	129	14	80	3	400	
Prod. acc.	47.7%	65.1%	92.9%	55.0%			
Overall accuracy: 56.4%							

Table A.2: Confusion matrix. Accuracy assessment of land cover change in building footprints as improved by Fura (2013), 2004-2010

Location is built up in period:		
Data from maps	Data from sample	
	Both	Total
Both	38	38
None	4	4
Only 2004	1	1
Only 2010	11	11
Total	54	54
Overall accuracy: 71.7%		

cover change, both likely and unlikely, and to stable land uses; in this sense, the sample overestimates the general error).

However, and *thirdly*, the problem of the land cover change does not seem to have been completely solved by Fura. Specifically, 11 out of the 54 points - 20% - were built up in 2004 and this was not detected in Fura's final maps¹ (it must also be noted that, relative to the entire Lubigi and Nakulolongo catchments, the area of Fura is much more urbanized; thus, most of the locations classified as non built in the sample are not within it; a general assessment is, in consequence, difficult to distill from the data).

A.3 LAND COVER IMPROVEMENT REQUIREMENTS

From the discussion on data availability and accuracy, the main points of improvement have been identified:

- Roads, in particular tarmac roads, needed to be introduced into the final land cover data model (specifically, the results of Fura need to be extended to the southern Lubigi and Nakulolongo areas).
- Building footprints were underestimating the existing built up for the 2004 year.

¹Fura's land cover map was reclassified, such that building footprints were considered built up and other land cover classes were all combined into the non built category for the assessment.

Initial explorations aimed to use ancillary (e.g. distance to roads, location of building footprints), transformed (principal components) and multi-temporal (both periods simultaneously) data to ease the classification of land cover for both years by pooling information from other periods. However, all the produced results included considerable errors. In addition, attempts to use supervised classification did not prove practicable, especially due to the diversity of spectral characteristics of buildings.

As a consequence, two methodological decisions were taken: (1) To create two land cover maps, each one for a separate year. (2) To use unsupervised classification methods in deriving the land cover maps. In addition, from the initial explorations, it was concluded that the building footprint information could be advantageously leveraged to inform the land cover classification maps.

The minimal mapping unit has been set at $2.25m^2$ (a $1.5m$ sided square). The original intent was to use a $1m^2$ unit but, on inspection of the imagery, it proved unfeasible to distinguish meaningfully such small elements, even at the most detailed scale (around 1:600). Further, using a pixel smaller than $1.5m$ sided cells introduced problems with the computability of the data (data sets became too large to be handled by the available hardware). Thus, the imagery was resized to $1.5m$ cells by estimating the average spectral value for a 3×3 cell window and subsequently resizing, with a nearest neighbor algorithm, to $1.5m$ size of cell.

A.4 LAND COVER MAP DERIVATION

The produced land cover aims to improve existing land cover maps in three ways: first, as noted, to better detect buildings. Second, to improve the contrast between bare soil and other land covers. Finally, the land cover classification was extended southwards, to cover the entire study region defined for this project.

To achieve these objectives, a five class (built up, vegetation, bare soil, shadow and tarmac) system was defined. Input data included the three visible bands of each year mosaic, resized as described in the preceding section. Pixels covered by detected existing building footprints were set to value 0, ensuring they would be classified together in a single class - easily interpretable as impervious land cover.

A maximum likelihood unsupervised classification algorithm was applied to separate imagery into nine classes, which were subsequently sorted into one of the four land cover classes (using the ArcGIS 10.2). Classes showed considerable confusion between the same two categories (or more) were reclassified as a mask; this mask was then applied to the original imagery and the unsupervised classification was then repeated. This procedure was successively applied until all resulting categories were unambiguously separated into one of the five land cover categories. The general procedure is schematized in figure A.2.

The resulting preliminary land cover was improved by: (1) digitizing elements from two additional land categories, water and no data (cloud cover and its shadow), (2) reclassifying the shadow category as vegetation: most of the cells of this category were, indeed, clearly dark vegetation or shadows of large trees (and only a minority, shadows of buildings), so the error produced by this simplification was judged to be small, (3) adding the missing building footprints to the 2004 vector data set (by visual inspection of the 2004 imagery, at a 1:1000 scale approximately; when a missing building was detected, it was copied from the 2010 vector data set, if existing and correct, and digitized if otherwise) and setting these new areas to the built up category; this process was only performed for cells *not* identified by the classification algorithm as built up.

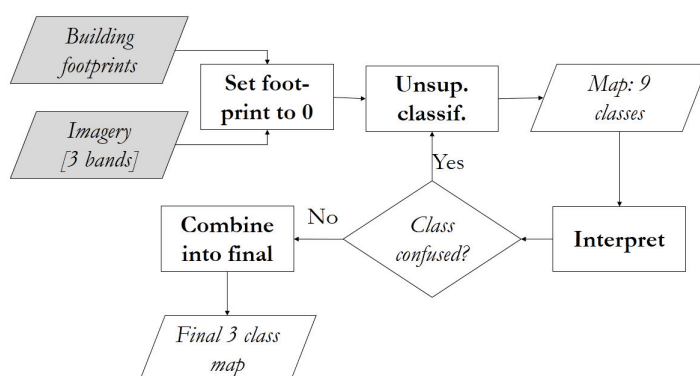


Figure A.2: Land cover map derivation procedure for the study area, 2004

Final results for the 2004 land cover are shown in figure A.3 (an equivalent map was also produced for 2010).

The general patterns seem to coincide with both existing land use maps and, broadly, with a visual assessment of the 2004 imagery. At a more detailed scale (see zoomed in areas reported in figure A.3), it can be seen that some, but not all, built up areas that were not detected in the building footprints feature, are being identified in the land cover map. Errors - particularly built up locations being classified as bare soil - remain but the problem seems to have substantially mitigated. A full comparative assessment of the overall accuracy is reported in subsection A.5.

A.5 COMPARATIVE ACCURACY ASSESSMENT OF LAND COVER DATA MODELS

Overall accuracy was evaluated for each of the three main land cover products: building footprints, Fura's land cover map and the land cover map produced for this study. A set of 400 random points, covering the entire study area, were classified by the visual assessment of the original imagery of each year. Of these points, 130 points were within the area analyzed in Fura (2013).

Overall accuracy was assessed as percent of correctly classified and also using the Kappa statistic, as programmed in the ArcGIS ModelBuilder Model developed by Reams (2010) (the Kappa statistic measures the improvements introduced by the classification, relative to a random assignment; which is to say, how much better is the land cover map than a random raster with the same categories).

Land cover maps of both periods (2004 and 2010) were separately assessed. For the Fura and the developed land cover maps, two versions were assessed: an aggregate, two category classification (built up and non built) and a full, five category classification system (for Fura's land cover map, the categories earth road and gravel were taken as bare soil). Points classified as no data, either in the sample or in the land cover maps, were ignored. Table A.3 summarizes the results.

As can be seen from table A.3, when considering the built up/non built land cover only, the developed land cover maps for both years represent a substantial improvement over both the original building footprints vector data set and the maps by Fura (2013). Overall accuracies are greater for the developed map by around 5% for 2004; for 2010, there is an improvement over Fura's result (overall accuracy of over 87% vs. 78% in Fura's map) and a very small, likely insignificant, 0.2% improvement with respect to the original building footprint. Kappa statistics of all two-class land

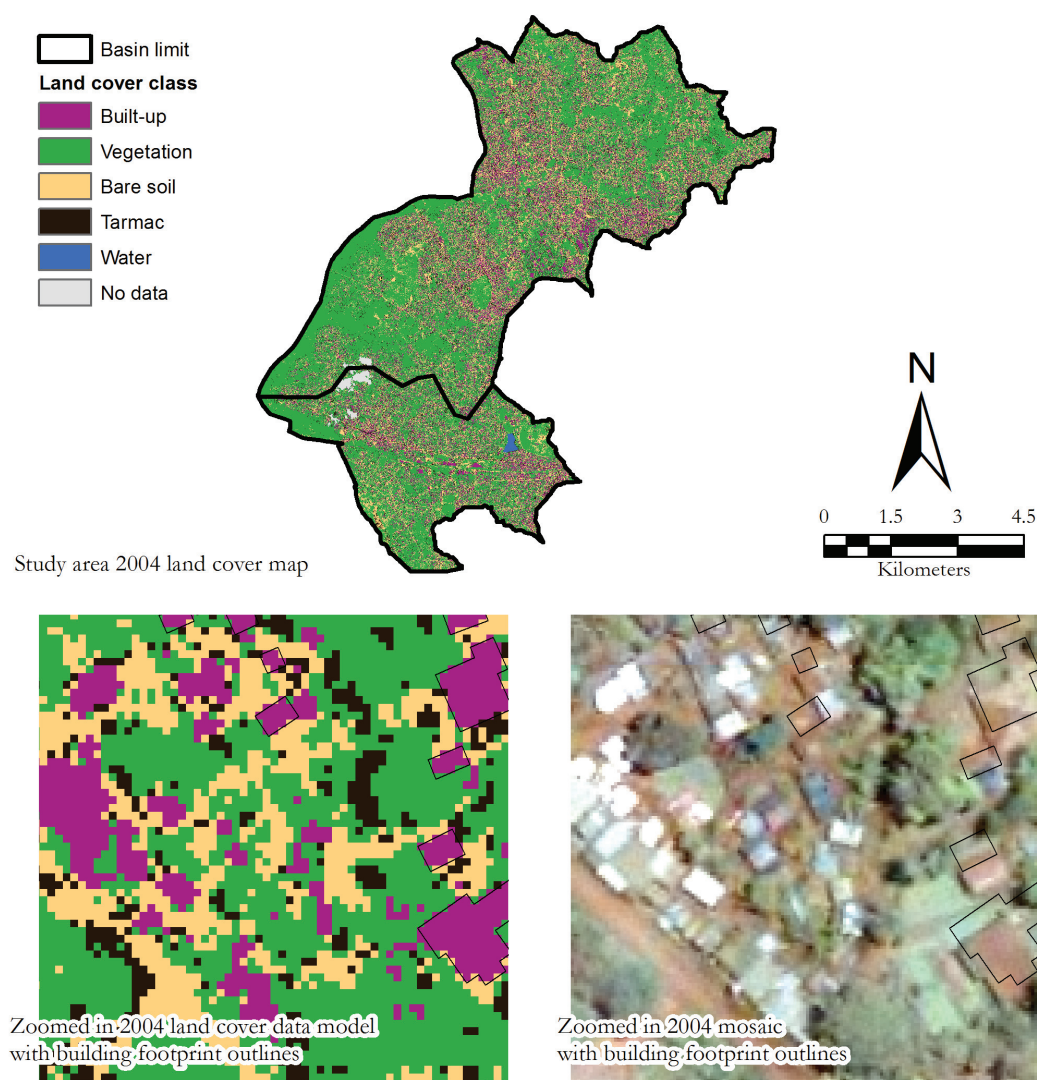


Figure A.3: Land cover map of the study area, 2004

Table A.3: Accuracy assessment of land cover data models. Overall accuracy and Kappa statistic

Data model	2004		2010		Sample size
	Overall acc.	Kappa	Overall acc.	Kappa	
Building footprints	84.9%	64.5%	87.3%	65.1%	400
Fura (2 categories)	84.4%	61.1%	78.3%	64.5%	130
Fura (5 categories)	72.3%	51.1%	73.8%	58.3%	130
Final map (2 categories)	88.9%	53.7%	87.5%	65.7%	400
Final map (5 categories)	69.8%	50.3%	74.0%	62.4%	400

cover products seem to be equivalent (all within the 60-66% range), except for the developed 2004 final map - which shows a substantially lower Kappa value (53.7%).

When considering the five category maps of Fura and the developed land cover map (see table A.3), overall accuracies and Kappa statistics are essentially equal. The developed map has a slightly lower

accuracy and Kappa for 2004, possibly because Fura's area was more urbanized and, generally, built up land cover is more easily detected. For 2010, both the overall accuracy percent and the Kappa of the developed map are slightly larger than Fura's results.

In general, the developed maps seem to be at least equivalent in quality to both the building footprints and the model developed by Fura, with the advantage of including more categories (with respect to the former) and a greater extent (with respect to the latter).

A.6 EXTENSION OF 2010 LAND COVER MAP

The Upper Lubigi subcatchment that was defined for simulation in the openLISEM model extended beyond the limits of the Kampala Capital City Authority's (KCCA) jurisdiction (see figure A.4). Because of this, no data was available for building footprints and roads in this area.

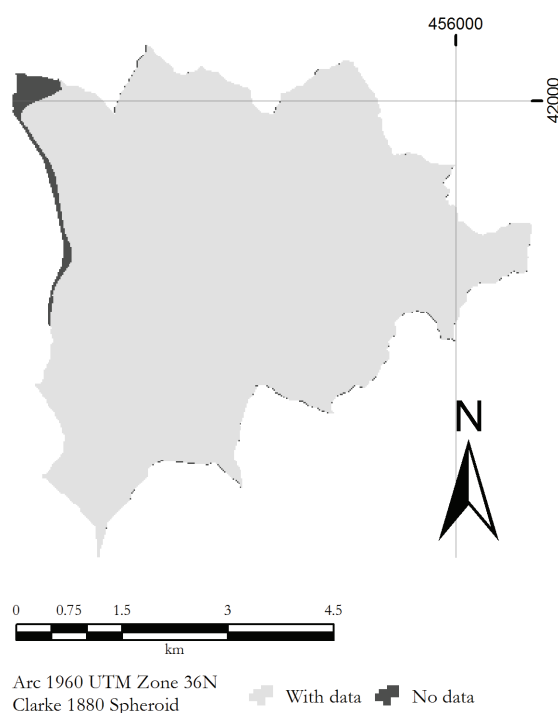


Figure A.4: Area for which 2010 land cover map was extended

Model calibration was performed without accounting for this difference, i.e., with all percentages of this excluded area set to 0% for all land cover categories. However, for the development of scenarios, land cover and road maps were extended to include this section. The following procedure was followed:

- Roads were digitized as centerlines, at a scale of 1:1000. The width of each segment was assigned by visual inspection and use of the measurement tools of ArcGIS 10.1. Only segments connecting with the network, whether within or beyond the study area, were digitized (as opposed to disconnected footpath segments).
- Building complexes (polygons enclosing several built up structures) were digitized.

- The original 2010 imagery was resampled to 1.5 m for the extent of the Upper Lubigi sub-catchment, with the buildings set to 0 value (including the newly digitized buildings).
- A supervised classification, using as a sample the derived map for 2010, was created.
- The final land cover map was compiled: the newly classified area was added to the original land cover map.

Appendix B

Generation of accessibility spatial index

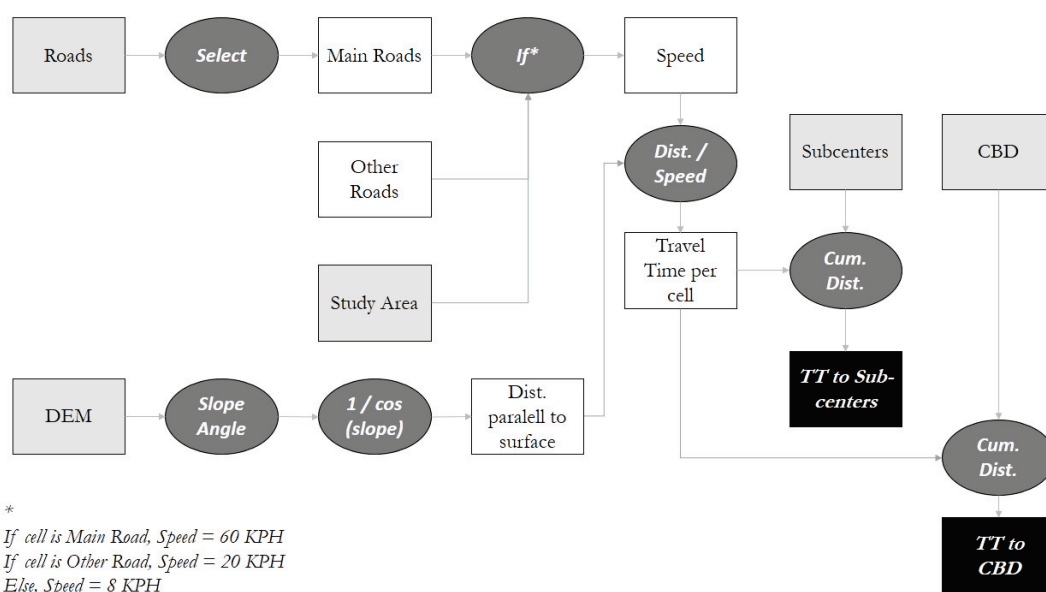


Figure B.1: Flow chart for the derivation of travel time to CBD and subcenters

Appendix C

Analysis of Land Cover Relationships of Built Up and Bare Soil Percentages, Neighborhood Effect and Other Inputs for Modeling

This section summarizes an exploratory analysis of the relationships between built up land cover change, bare soil land cover and a neighborhood value - defined as the average of built percentage up in the initial year within a moving window - for the part of the Lubigi catchment located within Kampala. It was performed on preliminary data; thus, the extent of the analyzed cells includes the mid and upper Lubigi (as opposed to the final calibration, which was done only for the upper Lubigi subcatchment). Additionally, unlike the final analysis, the square cells used as records (units) in the data base analyzed in this appendix had a length of 10 m.

Three general conclusions were inferred from this analysis and, later, informed the model development process. (1) A spatial statistical model, regardless of the amount of information used in it, cannot explain the amount of land cover change into built up land cover; it can, however, explain the locations where it occurs (which is essentially what non-linear models based on Logit econometrics do, such as the works by Fura, Abebe or Vermeiren and colleagues). (2) While a relationship is expected between bare soil and built up land cover, it is not simple. Bare soil land cover indicates several possible land uses or dynamics, only one of which is conclusively linked to urban growth. Further study of this issue is required. (3) The smaller the size of the moving window, used to estimate the neighborhood effect, the better its capacity to predict the location of built up land cover (in this appendix, the best size was a 3 cell, i.e., 30 m, moving window, the smallest possible; Fura, using a more detailed scale, adopted a 7 cell window of cells of 2.5 m).

C.1 SIMULATION OF BUILT UP LAND COVER DEMAND PATTERNS

Table C.1 and in figure C.1 summarize the descriptive statistical analysis of the cells exhibiting built up percent increase (growth). The raw percentages of growth show a skewed distribution but their logistic transformation does follow, broadly, a normal distribution.

An attempt was made to link this growth percentage to potential explanatory factors. The reasoning behind this is thus: if a cell has high development potential, which has not been realized, then this cell is the most likely to change in the short run. Evidently, this argument does not account for particular factors blocking development at a specific location. But the cells subject to this kind of restrictions should be, in principle, the exception rather than the rule.

Table C.2 shows the goodness of fit measures for a regression model, estimated using Ordinary Least Squares, attempting to explain the logistic transformation of change in the percent of built up land cover. Unlike the model originally used to compute demand in the preceding section, these

Table C.1: Descriptive statistics of built up percentage growth

Parameter	Built up %	Logistic transf. of built up % 1/
Mean	0.2976	-1.2328
Standard deviation	0.2382	1.7605
Skewness	0.8271	0.4356
Kurtosis 2/	-0.0868	5.1695
Minimum	0.0000	-9.2104
25th percentile	0.0982	-2.2161
Median	0.2431	-1.1357
75th percentile	0.4483	-0.2071
Maximum	1.0000	9.2005

1/ The logistic transformation was applied to all data cells, including negative and zero growth. To avoid losing information from the latter, a fraction of 0.001 was added to all percentages before applying the transformation. This correction does not affect the statistics in this table

2/ Neutral element of Kurtosis is 0.

results only included cells that actually grew in terms of their built up land cover (i.e., only positive values). It is very evident, from these results, that the factors selected to explain allocation (*where* development occurs) clearly are not determining *how much* development has happened. Adjusted R^2 values are all less than 2% and, what is more striking, they hardly improve as more elements are added to the models. Similar judgments can be made of alternate comparison measures: there is hardly any improvement in the RMSE and AIC when comparing between models.

Table C.2: Comparison of alternative explanatory demand models of logistic transformation of built up land cover growth

Model	$\hat{R}^2$	RMSE	AIC
Intercept	-	1.760	705683.9
Neighborhood	0.0148	1.747	703028.9
Neighborhood + Random	0.0148	1.747	703029.5
Neighborhood + Random + Others	0.0182	1.744	702414.9
2nd order polynomial Neighborhood + Random + Others	0.0184	1.744	702387.8

The alternative to an explanatory demand model is to assume that development is more or less random and its logistic transformation follows a normal distribution. To this end, random values between 0 and 1 were assigned to each cell. These were taken to be the associated probability of a standard normal value, z . The computed z was then re-scaled, using the mean and standard deviation of the logistic transformation of the built up land cover percent change (of cells showing growth only). The simulated growth was then assumed to be the inverse logistic transformation of this last value.

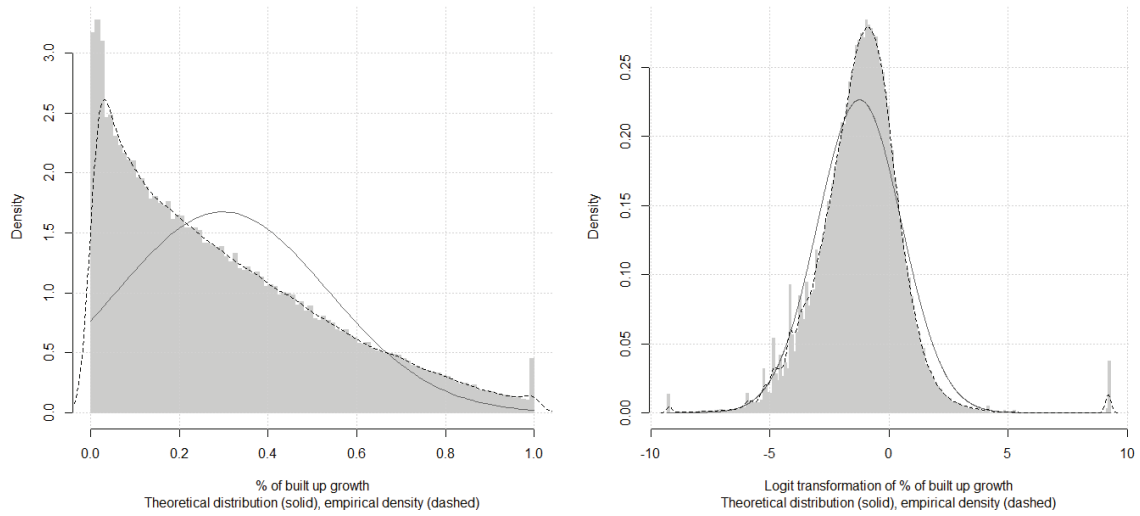


Figure C.1: Histograms. Percentage of built up and its logistic transformation for cells in which built up percentages increased, Lubigi, 2004-2010

C.2 NEIGHBORHOOD EFFECT: DEFINITIONS OF QUANTIFICATION AND INTERACTION

It should be expected, both from the theory (see, for example, the work of Tobler, 1979) and from empirical evidence for Kampala (particularly, the results of Fura, 2013), that land cover change into built up is explained to a large degree by the surrounding land cover: if an undeveloped locations is in the midst of a largely developed area, it is very likely that this location will be among the first to change into built up.

There are, however, two caveats: first, if an undeveloped location exists in a highly urbanized area, it likely should have been already urbanized; thus, the undeveloped state may very well be an indication of a particular condition that prevents that location from being developed. Second, when considering the spatially aggregated version of the land cover map, a cell with a very high percentage of built up land cover will have little available area for development - a condition consistent with a high value of built up land cover in its neighborhood -, so the land cover change *into built up* may be small despite this location being surrounded by built up land cover.

This second argument suggests that the relationship between the neighborhood (average built up percent of the surrounding cells) of the initial year, and the land cover change percent into built up have a concave relationship. If the neighborhood of the cell equals 100%, by definition there will be no change into built up because all the area is occupied. On the other hand, if the neighborhood percent equals 0%, there is little chance of the land cover change into built up occurring, as these are isolated locations. The maximum change into built up corresponds, therefore, to a neighborhood value located between 0 and 100%¹.

Figure C.2, the scatter plot of the change in percent of built up (from 2004-2010) vs. percent of built up within the neighborhood of the cell in 2004 for Lubigi. The neighborhood is the result of averaging the built up land cover percent in 2004 within a square moving window (of size varying

¹This is the same argument behind the famous Laffer curve, which in Economics describes the relationship between taxable income (government revenue from taxes) and the level of taxation.

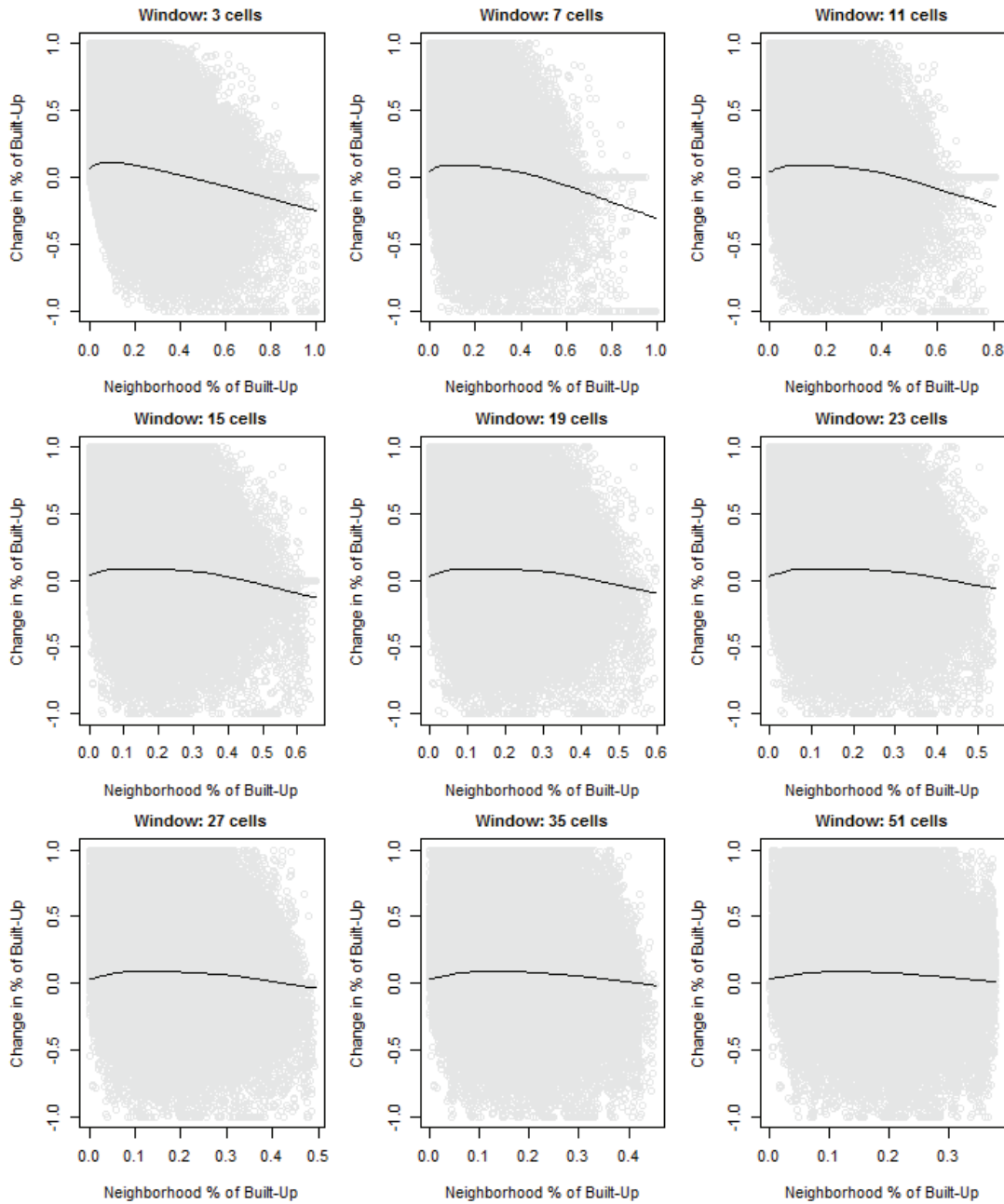


Figure C.2: Change in the built up Percentage 2004-2010 vs. Neighborhood built up Percentage 2004. Scatter-plot and smoothed trend (LOESS with bandwidth = 1/2)

from 3 to 51 cells, which is to say, 30 to 510 m). The total number of cells within Lubigi exceeds the 400 000. As can be seen in figure C.2, there is no evident trend within such a large data set. To visualize the trend, a locally weighted scatterplot smoothing trend line was estimated using the *loess.smoothing()* functionality in R (R Core Team, 2013) is included. As can be seen, the empirical relationship does resemble a concave parabola - in keeping with the discussed rationale. Also in figure C.2, the changes in the parabola as the neighborhood's window size increases, are also evident: for the lower values, the vertex is more clearly defined; as the neighborhood size increases,

the trend line tends to flatten, towards a horizontal line intersecting the y axis at (or near to) 0.

The question now is, which is the size of the moving window that should be selected to best explain urban growth? In a preceding section, proof was provided that the amount of change cannot be effectively predicted. But its location, which is to say, to classify cells according to whether they show increase, no change or decrease in the built up land cover percentage, is related to other factors - notably, the value of the neighborhood in the initial period analyzed.

Based on the discussed theoretical reasoning and empirical data, a relationship between change in built up percent and the neighborhood average built up land cover in the base year (as the determinant), was postulated (equation C.1). As can be seen, second degree polynomial form was assumed - and a negative sign for the β_3 value was expected, for the result to be a concave parabola.

$$\Delta BU_{04-10}^{xy} = \beta_1 + \beta_2 \cdot Neighb_{04}^{xy} + \beta_3 \cdot (Neighb_{04}^{xy})^2 \quad (C.1)$$

Equation C.1 was estimated using as determinant each one of the different neighborhood sizes available. Tables C.3, C.4 and C.5 show the $\hat{R}^2$, Root Mean Square Error (RMSE) and Akaike Information Criterion (AIC) for each equation, and for three variations: the original change in built up percentage (1) for the full data set, (2) excluding decreases, and (3) using a Logistic transformation (which also excludes decreases because negative percentages result in indeterminate values).

Generally, larger $\hat{R}^2$ and smaller RMSE, AIC values are desirable. As can be evidently seen, the smaller the window size, the better the results. Thus, of the selected windows, a 3×3 cell neighborhood is optimal (i.e., explains the most variation). This result is consistent with Fura (2013), who also chose the smallest possible window size as the optimal. Further, the Logistic transformation of the dependent variable increases, in all cases, the quality of the prediction. The full results

Table C.3: Comparison of predictive model of change in % of built up land cover, 2004-2010, as a function of average percentage of built up land cover in 2004 in moving window. Adjusted Coefficient of Determination

Explanatory variable	ΔBU (all)	ΔBU (≥ 0)	Logit ($\Delta BU \geq 0$)
Neighb. Window 3	0.0193	0.1029	0.3083
Neighb. Window 7	0.0100	0.0890	0.2417
Neighb. Window 11	0.0090	0.0834	0.2171
Neighb. Window 15	0.0083	0.0789	0.2016
Neighb. Window 19	0.0076	0.0746	0.1878
Neighb. Window 23	0.0070	0.0708	0.1760
Neighb. Window 27	0.0066	0.0675	0.1659
Neighb. Window 35	0.0057	0.0613	0.1488
Neighb. Window 51	0.0046	0.0511	0.1218

ΔBU : change in built up land cover percentage, 2004-2010

Logit: logistic transformation of ΔBU

To all percentages, a fraction of 0.001 was added to avoid null values in the logistic transformation.

of equation C.1, for the Logistic transformation is shown in equation C.2. Below each coefficient is indicated the t value for it - all are significant with $Prob. < 0.001$.

$$\begin{aligned} \text{Logit}(\Delta BU_{04-10}^{xy}) = & -10.878 + 50.114 \cdot Neighb_{04}^{xy} - 63.880 \cdot (Neighb_{04}^{xy})^2 \\ & (864.6) \quad (324.1) \quad (192.0) \end{aligned} \quad (C.2)$$

Table C.4: Comparison of predictive model of change in % of built up land cover, 2004-2010, as a function of average percentage of built up land cover in 2004 in moving window. Root Mean Square Error

Explanatory variable	ΔBU (all)	ΔBU (≥ 0)	Logit ($\Delta BU \geq 0$)
Neighb. Window 3	0.23091	0.19851	5.29566
Neighb. Window 7	0.23200	0.20005	5.54464
Neighb. Window 11	0.23212	0.20067	5.63389
Neighb. Window 15	0.23220	0.20115	5.68928
Neighb. Window 19	0.23228	0.20163	5.73808
Neighb. Window 23	0.23235	0.20203	5.77981
Neighb. Window 27	0.23239	0.20239	5.81506
Neighb. Window 35	0.23250	0.20306	5.87452
Neighb. Window 51	0.23263	0.20416	5.96667

ΔBU : change in built up land cover percentage, 2004-2010

Logit: logistic transformation of ΔBU

To all percentages, a fraction of 0.001 was added to avoid null values in the logistic transformation.

Table C.5: Comparison of predictive model of change in % of built up land cover, 2004-2010, as a function of average percentage of built up land cover in 2004 in moving window. Akaike Information Criterion

Explanatory variable	ΔBU (all)	ΔBU (≥ 0)	Logit ($\Delta BU \geq 0$)
Neighb. Window 3	-40 271.1	-134 678.6	2 099 348.0
Neighb. Window 7	-36 213.8	-129 416.2	2 130 605.0
Neighb. Window 11	-35 756.3	-127 333.7	2 141 468.0
Neighb. Window 15	-35 458.3	-125 677.9	2 148 124.0
Neighb. Window 19	-35 168.8	-124 085.6	2 153 935.0
Neighb. Window 23	-34 904.4	-122 712.4	2 158 864.0
Neighb. Window 27	-34 742.2	-121 505.1	2 163 001.0
Neighb. Window 35	-34 340.5	-119 252.9	2 169 922.0
Neighb. Window 51	-33 859.1	-115 573.9	2 180 511.0

ΔBU : change in built up land cover percentage, 2004-2010

Logit: logistic transformation of ΔBU

To all percentages, a fraction of 0.001 was added to avoid null values in the logistic transformation.

C.3 BARE SOIL AND BUILT UP LAND COVER

For bare soil, it should be expected that, as the percentage of built up land cover increases, so will the bare soil (at the expense of vegetation). More built up land cover implies greater human activity, in particular transport, near this development. Bare soil should increase because of people walking around the new development, trampling vegetation, or because of new (unpaved) roads. This same reasoning was adopted by Fura (2013, p. 64), who assumed the proportion of bare soil (relative to the original built up area) is maintained as new built up land cover is constructed.

The empirical relationship between the percentage of built up in the base year (2004) within the neighborhood of the cell and its percentage of bare soil change (2004-2010) is shown in figure C.3. Again, the sheer amount of points confuses the interpretation so a locally weighted scatterplot

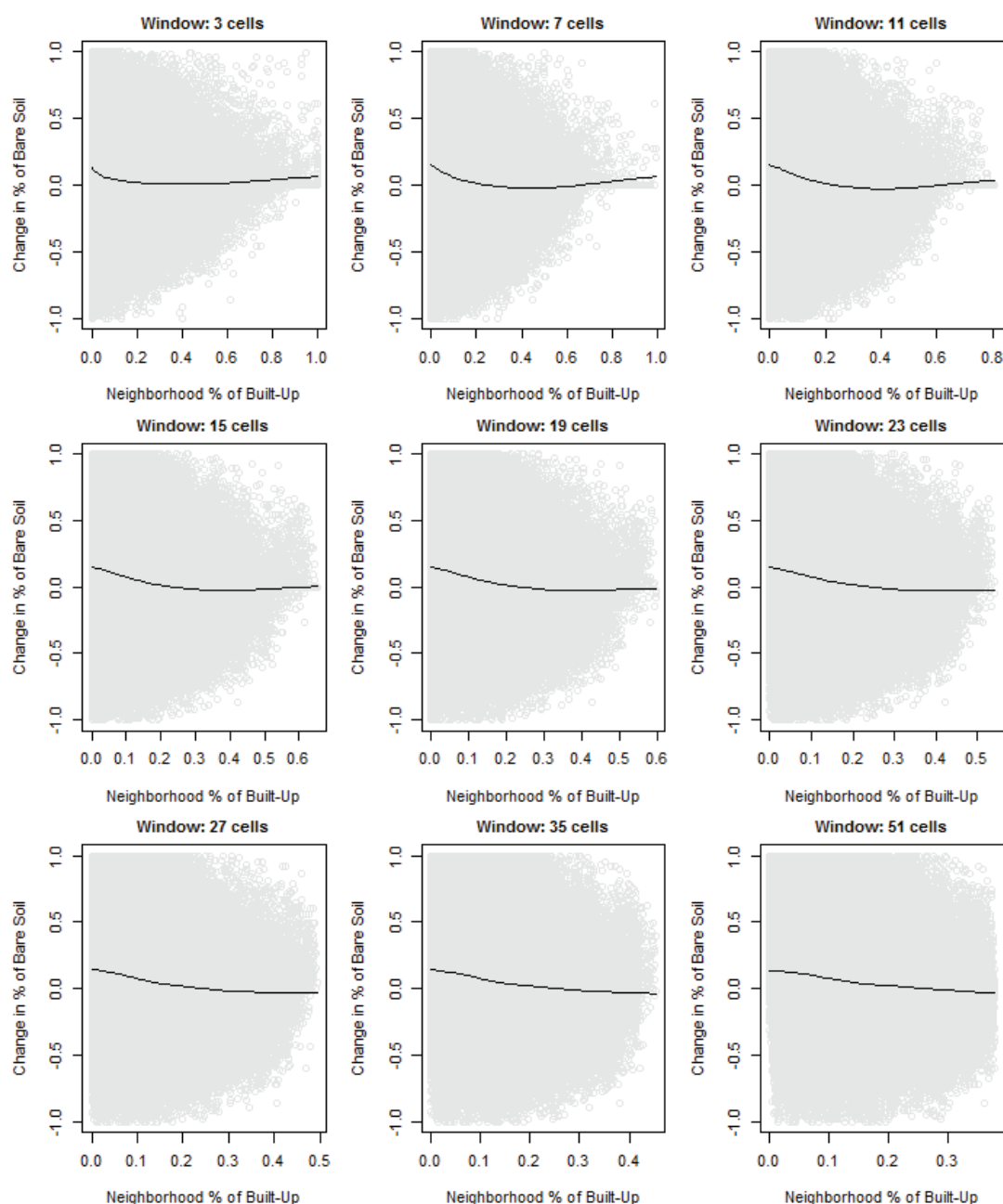


Figure C.3: Change in the Bare Soil Percentage 2004-2010 vs. Neighborhood built up Percentage 2004. Scatterplot and smoothed trend (LOESS with bandwidth = 1/2)

smoothing trend line was added. Unlike the relationship of built up, which was more clearly concave, the bare soil trend is not evident. For lower neighborhood values, there is generally a decreasing trend. But with most values in between, the relationship resembles a flat, linear horizontal at the level of 0 (which is to say, there is no average change related to the neighborhood average). As the neighborhood size increases, a linear, decreasing (with a small slope) trend line seems to emerge. The relationships for smaller window sizes show (a little) more variation.

The analysis of the relationship between bare soil and built up land cover, both in terms of change

and level, does not reveal any systematic variation.

The hypothesized mechanism, that bare soil increases with urban development, is clearly insufficient. At least three additional factors may be occurring as well: in developed areas, tress and grass has substituted bare soil in the backyards of mid-to-high income housing; as noted by Fura (2013), the season in which each of the base images (2004 and 2010) was taken, is different - specifically, the 2010 image is greener than the 2004 image; finally, the boundary between bare soil and vegetation is fuzzy: it is difficult to distinguish very degraded grass from bare soil proper, which may have led to some confusion in the land cover data models. To properly model this aspect of land cover change in Kampala, a deeper exploration of the variation of vegetation and bare soil fractions in each cell would be required. This analysis could not be performed due to time limitations inherent to the project.

In the absence of a clear mechanism describing the aggregate relationship between bare soil and built up land cover, the assumption of Fura (2013) - that the ratio of bare soil and built up percentages remains constant across time periods - may be partially maintained. This assumption is conservative, in the sense that vegetation exhibits a higher infiltration rate than highly compacted urban bare soil, so the rule would (in a worst case scenario) lead to an overestimation of potential flooding effects. This exaggeration of the impacts of flooding is inefficient: it may lead to restrictions on potentially safe land. But at least it would reduce the overall impact and increase quality of life, while simultaneously allowing, in part, for rare catastrophic events.

Appendix E

Statistical Assessment of Final Land Cover Model Results

Table E.1: Predicted built up land cover percentage, 2010, using multiple factor models with varying weights. Comparison of predictions and land cover map of calibrated model

Model	Correlation	KS	Kappa	Morans's I	Sum of BU%
Land cover map 2010	-	-	-	0.4517 $z = 237.8$	15950
1. TT to CBD weight = 0.5	0.5450	0.1098 $Pr. < 0.001$	27.8%	0.4140 $z = 207.1$	14749
2. Wetland weight = 2	0.5536	0.1153 $Pr. < 0.001$	27.3%	0.4320 $z = 201.3$	15029
3. TT to CBD weight = 0.5 and wet. = 2	0.5494	0.1087 $Pr. < 0.001$	27.7%	0.4165 $z = 226.8$	15116
4. TT to CBD & non-veg. = 0.5, wet. = 2	0.5412	0.1149 $Pr. < 0.001$	27.6%	0.4012 $z = 230.3$	14327

Correlation: Pearson Correlation Coefficient.

KS: Kolmogorov-Smirnov two-tailed test; null hypothesis: both samples come from the same probability distribution; all null hypothesis rejected with Prob. < 0.001 .

Kappa: Cells reclassified into five categories (0-20%, 20-40%, 40-60%, 60-80%, 80-100%); Cohen's Kappa between the land cover map and the prediction, was estimated for this reclassification, using the program of Reams (2010).

Moran's I: estimated using first order adjacency (including corners), row standardized

Appendix F

Performance of CA Model

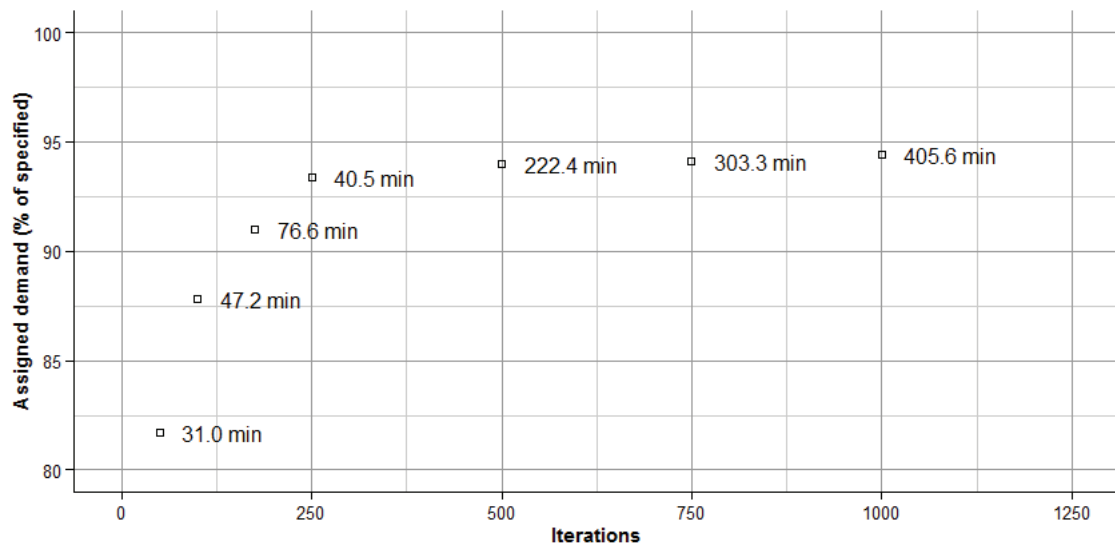


Figure F.1: Model Performance: Number of Iterations in Allocation Routine vs. Percentage of Exogenous Demand Assigned for Final Calibrated Model

Appendix G

Simulated land cover for 2020 scenarios

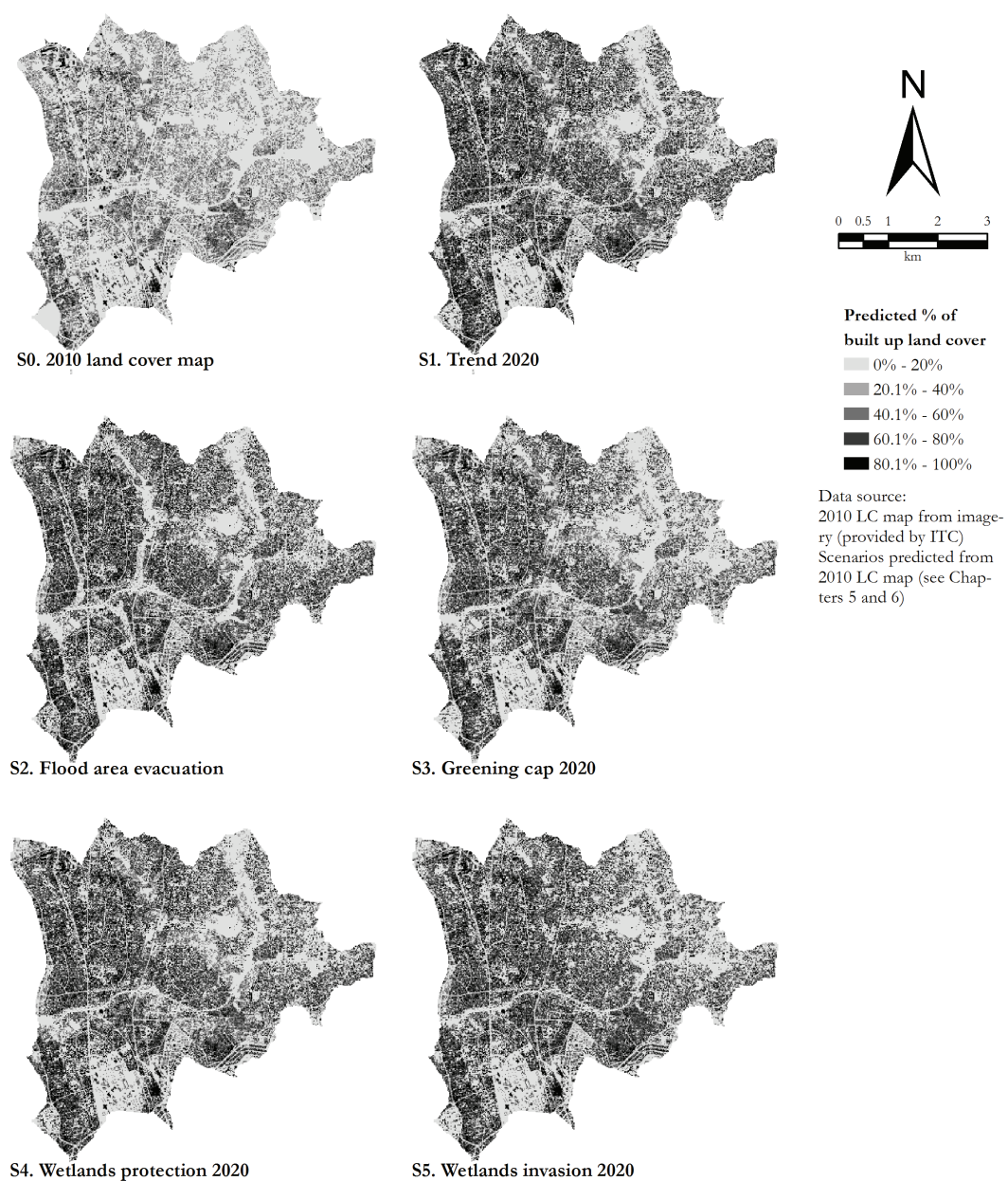


Figure G.1: Predicted land cover maps for specified scenarios, 2020

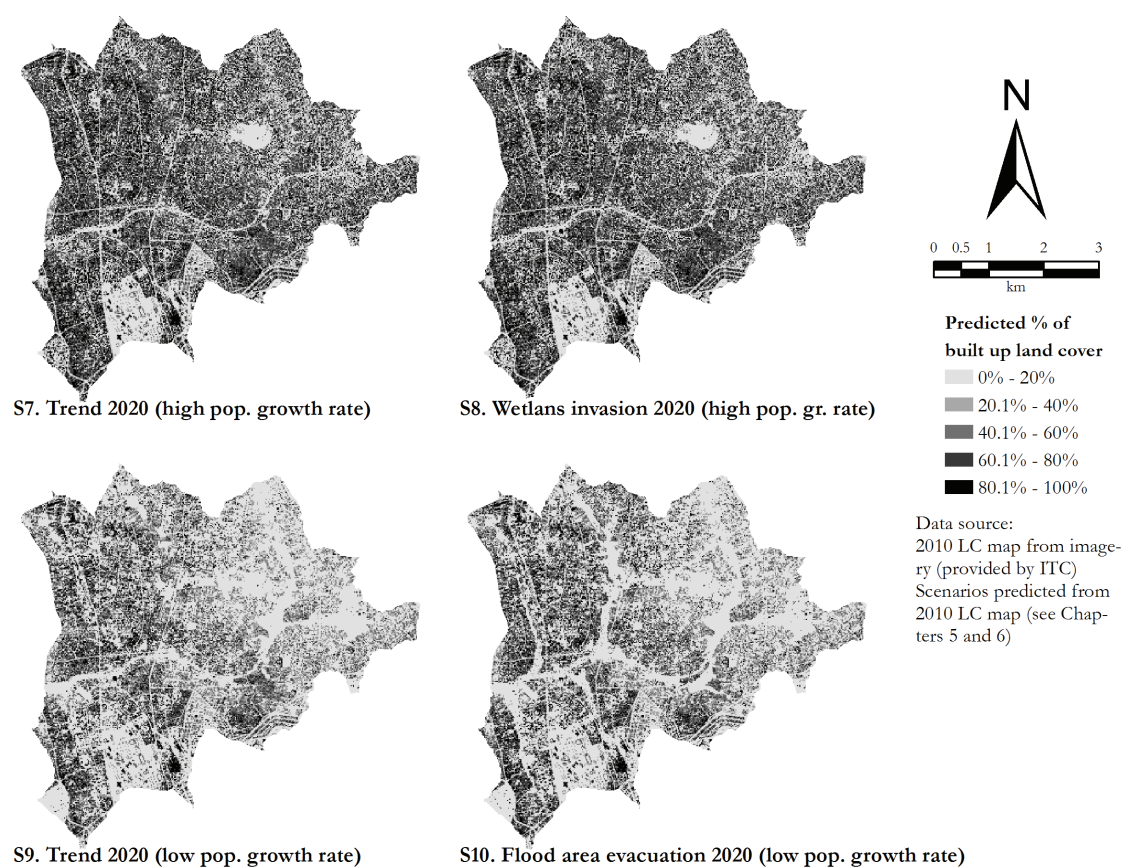


Figure G.2: Predicted land cover maps for specified scenarios, 2020