Université catholique de Louvain
Faculté des Sciences, École de Géographie
Center for Operations Research and Econometrics

Spatial bias in LUTI models

Doctoral Dissertation presented by

Jonathan Jones

in fulfillment of the requirements
for the degree of Doctor in Sciences

Thesis committee:

Prof.	Dominique Peeters (Supervisor)	Université catholique de Louvain
Prof.	Isabelle Thomas (Supervisor)	Université catholique de Louvain
Prof.	Marie-Laurence De Keersmaecker (Chair)	Université catholique de Louvain
Prof.	Geoffrey Caruso	Université du Luxembourg
Dr.	Philippe Gerber	Luxembourg Institute for Socio - Economic Research
Prof.	Bart Jourquin	Université catholique de Louvain

Louvain-la-Neuve, Belgium
June 2016

Cette thèse n'aurait, au sens propre, pas pu voir le jour sans mes deux promoteurs, Dominique Peeters et Isabelle Thomas. Laissé à lui-même, l'auteur de ces lignes ne se serait probablement pas orienté vers la recherche, mais "le chef" et "la patronne" ont estimé, aux alentours du mois de juin 2009, que mon profil correspondait à celui recherché pour prendre en charge un poste d'assistant à l'école de géographie de l'UCLouvain. Mes premiers remerciements vont donc naturellement vers eux (et peut-être aussi à mon prédécesseur, Alain Pholo Bala, pour avoir terminé sa propre thèse juste à temps). Merci pour m'avoir fait confiance à l'époque, mais surtout pour l'encadrement offert au cours de mon doctorat. Encadrement incluant (mais non limité à) de nombreux bons conseils sur l'orientation à prendre en début de thèse, les avis éclairés sur les principaux choix méthodologiques, une disponibilité sans faille... Et un trajet Louvain-La-Neuve - Dourdan mémorable.

Bien que mon sujet de recherche ait parfois fait de moi un "outlier", j'ai également bénéficié de conseils avisés de la part de nombreuses autres personnes. En premier lieu, les partenaires du projet SustainCity. Diverses analyses conduites dans cette thèse n'auraient pas été possible sans leur travail. Je dois ici remercier en particulier Ricardo Hurtubia, Dimitrios Efthymiou, Inès Cabrita, et Thomas Nicolaï ainsi que toute l'équipe "Géo - en - CORE" : Adeline, Arnaud, Gaétan, Justin, et Sonia.

Malgré une "distance - decay" plus importante, le fait de pouvoir s'intégrer à un petit réseau d'autres doctorants (Summer School de l'European Regional Science Association et participation à trois éditions de l'European Colloquium on Theoretical and Quantitative Geography) fut incontestablement un plus. Luke Webber m'a apporté une aide nécessaire en fin de rédaction, pour corriger mes fautes d'anglais. Toutes les erreurs restantes demeurent bien sûr de ma responsabilité. Enfin, je voudrais remercier les membres de mon jury (Geoffrey Caruso, Philippe Gerber, et Bart Jourquin) pour le regard critique mais globalement bienveillant qu'ils ont bien voulu porter sur mon travail.

Mais tout comme le cyclisme, une thèse est affaire d'endurance. Pour éviter la saturation, ou surmonter les inévitables phases de découragement, je suis convaincu que tout doctorant a besoin d'un cadre de travail agréable, ce qui n'a jamais manqué d'être le cas pour moi. Les responsables des nombreux fous rires lors de la pause café, des temps de midi à rallonge à prendre le soleil devant le Mercator, et des (trop) nombreuses Triple Karmeliet après une longue journée de travail se reconnaitront. Qu'ils sachent que ce travail n'aurait sans doute pas abouti sans eux.

Pendant mes mandats d'assistant, j'ai eu la chance de découvrir et - j'espère - d'aider à former sept générations successives d'étudiants. Du BAC1 jouette (repose ce crayon de couleur !) au M2 curieux (oui, mais pourquoi ?) en passant par le BAC2 timide (vous rejetez ou ne rejetez pas l'hypothèse nulle ?). Malgré le caractère rébarbatif de certaines des matières que j'ai eus à leur enseigner, l'encadrement des séances de travaux pratiques a souvent constitué une respiration bienvenue dans la routine de la thèse, et fut toujours un plaisir. Merci à vous tous pour l'enthousiasme et la curiosité dont vous avez fait preuve. Cette expérience d'enseignement a été immensément enrichissante pour moi, et reste le meilleur souvenir de mon parcours à l'UCLouvain.

Au cours de ma thèse, j'ai aussi pu compter sur le soutien de ma famille et de nombreux amis. A nouveau, les anciens de Basse - Wavre, les vétérans de la 21ième SV Limal, et tous les autres, se reconnaitront. Ils ont réussi à passer outre les trop nombreuses références à Phd Comics et autres private jokes sur la vie de doctorant pour me conserver leur amitié. Je leur en suis reconnaissant.

List of publications

The chapters of this thesis are based on the following publications (current and future):

- **Chapter 3:** Jones J, Peeters D, Thomas I, 2015. Is cities delineation a pre requisite for urban modelling? The example of land price determinants in Brussels. Cybergeo: revue européenne de géographie, `https://cybergeo.revues.org/26899`
- **Chapter 4:** Jones J, Thomas I, Peeters D, 2015. Forecasting employment location choices by Discrete Choice Models: A sensitivity analysis to scale and implications for LUTI models. REGION, Volume 2, Number 1, 2015, 67–93
- **Chapter 5 (mono centric case study):** Jones J, Peeters D, Thomas I, 2013. Does size matters? LUTI models, policy evaluations, and the scale effect. Proceeding of the 18^{th} European colloquium of theoretical and quantitative geography, Dourdan (France), 5 - 9 September
- **Chapter 5 (poly centric case study):** Thomas I, Caruso G, Jones J, Gerber P, Ongoing. Does city delineation really matter in LUTI models applications? To be submitted in Transport review
- **Chapter 6:** Jones J, Peeters D, Thomas I, Ongoing. Scale effect in a LUTI model of Brussels: challenges for policy evaluations. Submitted to the European Journal for Transportation and Infrastructure Research

Main acronyms

BCR .... Brussels - Capital Region

BSU .... Basic Spatial Units

CBD .... Central Business District

DCM .... Discrete Choice Models

EU .... European Union

LUTI .... Land Use and Transport Interactions models

MAUP .... Modifiable Areal Unit Problem

MPO .... Metropolitan Planning Organizations

US .... United States of America

UR .... Urban Region (of Brussels)

“Qu’est-ce au juste qu’une montagne? En donner quelque définition simple - l’ensemble des terres méditerranéennes au-dessus de 500 mètres par exemple - inutile précision. C’est de limites humaines, incertaines, malaisées à reporter sur la carte, qu’il doit être question"

Fernand Braudel (La Méditerranée et le monde méditteranéen à l’époque de Philippe II - La part du milieu, 1966; pp. 34)

“Quand on écrit un roman, on est Dieu le Père, parce qu’on crée le destin (...). Quand vous faites un film, vous êtes roi, parce que le destin est déjà là : il y a un scénario et vous ne pouvez pas vous égarer (...). Et quand vous êtes un documentariste, vous êtes un humble esclave qui marche derrière et ramasse les traces qu’ils ont laissées derrière eux."

Pierre Schoendoerffer (entretien à *Marianne*, 26 février 2007)

Part I

Introduction and state-of-the-art

CHAPTER 1

General introduction

1.1 Research focus and motivations

Cities are a complex phenomenon, involving a large variety of demographic, economic, and environmental processes. Nowadays, they also accommodate most human beings. The world's urban population surpassed the rural population around 2005 (UN, 2014), and continues to grow since reaching 75% in the European Union (EEA, 2013). There is, therefore, a need to understand how cities function, not only to reduce the social or environmental impact of sub urbanisation (see e.g. Kahn, 2000; Glaeser and Kahn, 2003; and Glaeser and Kahn, 2008), but simply because it is where most people live. As other research topics of social and human sciences, metropolitan areas cannot, however, be studied through direct experiments. Hence, studying cities has always led to the development of "virtual laboratories", i.e. theoretical and empirical models. Let us simply cite the three classical descriptive model of urban geography: the concentric zone model of Burgess (1925), the sector theory of Hoyt (1939) and Hoyt (1965), and the multiple nuclei model of Harris and Ullman (1945); or two of the most famous urban economics models, the urban bid-rent model of Alonso (1964) and the equilibrium non-mono centric model of Fujita and Ogawa (1982).

All of these models were initially one-dimensional or assuming a featureless isotropic landscape. Although it makes sense in explanative models studying the structure of cities resulting from the influence of one or a few processes, this is obviously not the case for predictive model forecasting the future land use pattern of metropolitan areas. This latter category encompasses two frameworks: Cellular Automata (CA, see White and Engelen, 1993; Batty, 2005; White et al., 2015) and Land Use and Transport Interactions (LUTI) models (see Wegener, 2004; Hunt et al., 2005).

A reasonably realistic representation of space is expected from these models, as well as an assessment of their sensitivity to the Modifiable Areal Unit Problem (MAUP), i.e. to the fact that identical individual data yield different statistical results when aggregated in varying ways. Such works can be found for CA models, showing an influence of the size of the pixels on the predicted land use pattern (e.g. Jenerette and Wu, 2001; Jantz and Goetz, 2005; Menard and Marceau, 2005; or Kocabas and Dragicevic, 2006). Different studies also exist for aggregated (four-stage) transport models, studying the variations of the forecasted trips' length and frequency induced by changes in the size or shape of the traffic analysis zones (e.g. Chang et al., 2002; Zhang and Kukadia, 2005; and Viegas et al., 2009). On the contrary, for LUTI models, no comparable sensitivity analysis could be found.

This thesis focuses on this later gap, and more precisely on the land use side of dynamic LUTI models which constitute state-of-the-art modelling in that field. We will examine both the sensitivity to spatial bias of the econometric methods used to forecast the evolution of the study area (the *behaviour* of the model) and of the final situation predicted (the *outputs*). This spatial bias is induced by the two main spatial choices that modellers face when using a LUTI model, the *spatial extent* and the *spatial resolution*. Our general objective is the following:

General objective *Assess the sensitivity of the behaviour and outputs of LUTI models to the variations of either the spatial extent or the spatial resolution of the model.*

The *spatial extent* refers here to the choice of the boundaries of the study area. It relates to the challenge of identifying the limit between urban and rural areas, that do not correspond anymore to a straight line but (as the Mediterranean mountains described by Braudel) to a fuzzy area determined by human activities (Cavailhès et al., 2004; Caruso et al., 2007). The *spatial resolution* is more technical. It designates the size of the minimal areal units existing in the model. For various practical reasons, most LUTI models remain spatially aggregated and the size of the areal units chosen as zoning system may thus influence the model system.

The relevance of this research derives from the importance of LUTI models in land-use planning and policy evaluation. LUTI models have enjoyed a growing interest since the beginning of the 1990s, for both scientists and planners, thanks to their ability to forecast future urban pattern and to assess the influence of environmental policies (Rodrigue et al., 2009). Operational applications of such models originally appeared in the US (Rosenbaum and Koenig, 1997; Bartholomew, 2007) Nowadays, several different LUTI models exist (see Hunt et al., 2005; Simmonds et al., 2013; or Wegener, 2014), that have been applied on numerous case studies, especially in the US and in Europe. For instance, the recent EU-funded project *SustainCity* (2009 – 2013) aimed to develop operational applications of a micro-simulation LUTI model for Brussels, Paris, and Zürich (see Bierlaire et al., 2015). LUTI models are even widely used for policy evaluation (see Badoe and Miller, 2000; Geurs and van Wee, 2004; and Bartholomew, 2007 for reviews). In the US, it is mandatory for metropolitan planning organisation to take into account the feedback effect of transport on land-use when applying for federal funding for transportation infrastructure improvements (Dowling, 2005; Waddell, 2011). They have, therefore, to rely on LUTI models as modelling framework when assessing the potential influence of such projects.

Our intention, in this thesis, is to demonstrate the importance of spatial choices for the robustness and goodness-of-fit of operational applications of LUTI models. The precise definitions of LUTI models, spatial extent, and resolution are given in section 1.2. Section 1.3 describes the main methodological choices and the outline of the thesis.

1.2 Terminology

1.2.1 Land-use and transport interaction model

The term "LUTI model" encompasses a large variety of frameworks (see chapter 2). The key factors, nevertheless, emerge clearly. For Wegener and Furst, 1999a, LUTI models must be *integrated* and *operational*. In a nearly identical way, Hunt et al. (2005) consider that such models have to be *integrated*, *comprehensive*, and *operational*. These three components mean that the model should (a) explicitly represent the links from transport to land use, and vice-versa. It should also (b) account for a complete range of spatial processes, especially the evolution of land use, through modelling the location choices of households and employment (comprehensiveness). Finally, (c) at least one application for policy analysis on a metropolitan region should exist. In a similar way, according to Wegener (2014), LUTI models *" explicitly model the two-way interaction between land use and transport to forecast the likely impacts of land-use policies (...) [or] transport infrastructure investments (...), for decision support in urban planning. That excludes transport models per se, which predict traffic patterns (...) and land-use models change models that predict likely land-use changes (...), as well as models that deal only with one urban subsystem, such has housing or business location"* (Wegener, 2014; pp. 38).

Therefore, LUTI models have to represent agents (at least households and employment) in some way (either by simple totals per zone or by a fully disaggregated representation), which excludes from the scope of LUTI models Cellular Automata where land use is only accounted for by a condition (e.g. rural versus urban). Two consequences emerge from this representation of agents. First, LUTI models generally rely on a zoning system composed of areal units of irregular size and shape (statistical or administrative units rather than grid cells). Secondly, LUTI models use econometric methods to predict the evolution of the quantity of agent in each zone (rather than on transition rules to predict changes of the state of the pixels as in a CA model). Finally, LUTI models are applied models, attempting to predict the future structure of the city rather than understanding the demographic or economic processes leading to this structure. Hence, to specify our framework, the definition of LUTI model that will be used in this thesis is the following:

Definition 1 *A land-use and transport interaction model (LUTI) is an applied model aiming at forecasting the evolution of a metropolitan area. It has to (a) integrate explicitly the two-way interactions between land-use and transport, (b) to implement a comprehensive set of spatial processes, including at least the evolution of land use and the location choices of both households and employment, and (c) to allow operational uses for policy evaluation. The zoning system consists in a set of areal units able to accommodate various activities (i.e. not defined by a particular condition as in CA model). LUTI models forecast the level of activities in each zone through time by using a sequence of sub models mainly based on econometric methods.*

Note that we will make a distinction between how a LUTI model works and the results that it produces. In this thesis, the *behaviour* of a LUTI model will designate the set of (econometric) methods by which it will forecast the evolution of the study area. This behaviour is specific to each LUTI model, although common features can be identified (see chapter 2). The *output* of a LUTI model is the situation predicted at the end of the simulation period. Its nature (e.g. number of inhabitants and jobs per municipality) may vary depending on the LUTI model used.

1.2.2 Spatial bias

Spatial bias is understood here as any modification in the representation of space that may affect either the *behaviour* or the *outputs* of a LUTI model. The conceptual background of these spatial bias are detailed in chapter 2. As indicated in section 1.1, we will focus here on two spatial choices made by the modellers.

Definition 2 *The spatial extent designates the size and composition of the study area on which a LUTI model is applied. The process by which a change* *in this spatial extent will - or will not - influence the behaviour and outputs of the model is referred to as the boundary effect.*

Definition 3 *The spatial resolution defines the zoning system used by a LUTI model, i.e. the size and shape of its basic spatial units. The potential sensitivity of the model to a change of its spatial resolution is called the scale effect.*

Note that the terms areal units and Basic Spatial Units (from now on BSU), although similar, convey different meaning. Basic Spatial Units (abbreviated BSU) will strictly be used here to designate the minimal spatial units existing in a LUTI model, i.e. its spatial resolution. The term areal units is to be understood in a broader sense as the result of a given zoning system, i.e. of a particular partition of space. The areal units for which a given variable is available can, for instance, be different from the BSU used in the model.

1.3 Methodological choices and outline of the thesis

Methodological choices have to be made to reduce our research questions to a feasible and consistent experiment plan, i.e. a sequence of analysis allowing an answer as complete as possible of our general objective while respecting the limited time available, as described by the outline of the thesis. Many of them emerged during the analyses. Hence, this section intends to present in an orderly fashion the "meta-choices" defining (1) which analyses will be conducted in the following chapters, and (2) why they were designed in that particular way.

1.3.1 Context of the thesis

The *SustainCity* project

This thesis was partially conducted under the framework of the EU-funded *SustainCity* project (2009 - 2013). The UCLouvain contributed to two of its work packages. The main task was to study the influence of spatial bias in LUTI models, which has led to this thesis. The second task was to help the development of the Brussels case study. For this latter task, the UCLouvain team has been responsible for the collection and processing of all data related to land use in this specific case. Note that the remaining data processing steps (especially the generation of the synthetic population) and the calibration of the model were under the responsibility of other stakeholders of the project. The *SustainCity* project constrained several methodological choices, in particular, the LUTI platform used. In this thesis, the zone version of the *UrbanSim* model (OPUS v 4.3) will be used in conjunction with *MATsim* as a representative LUTI model system.

The *UrbanSim* model only forecasts the evolution of land use (see chapter 2). The transport dimension is added by coupling it to an external transport model, here *MATsim*. As indicated, their use was determined in the *SustainCity* project proposal, but several reasons make the use of *UrbanSim* relevant. First, it is a disaggregated model with an individual representation of agents, and its internal principles are similar, although somewhat simpler, to those of other state-of-the-art LUTI models (e.g. IRPUD, Delta, and PECAS; see chapter 2 and chapter 4). In particular, *UrbanSim* relies on regression methods and discrete choice models to forecast, respectively, the evolution of real estate prices and agents' location choices. It would not have been possible to assess the sensitivity of different LUTI models to spatial bias in this thesis. Nevertheless, this focus on *UrbanSim* remains one of our main methodological weaknesses, and its implications will be discussed in chapter 7. Secondly, from a practical point-of-view, *UrbanSim* is a free and open-source software, and no acquisition or consultancy fees are therefore required for its use. The source code being public, no "black box" problem should appear, allowing reproducibility of the results. Both factors are, in theory, incentives for the use of *UrbanSim*. From a user point of view, however, this resulted in a lack of detailed documentation and technical support (limited to the online discussion forum). The learning curve is, therefore, very slow and consists mostly in trial and error. Hence, one often feels as if facing not a "black box" but rather a "black hole" when trying to develop operational application of *UrbanSim*.

Focus on land use

This thesis focuses on the land use side of the "Land Use and Transport Interactions" models. LUTI models (at least those distributed by private consultancy companies, see chapter 2) are rarely integrated models. They constitute rather a coupling between a land use model (e.g. *UrbanSim*) and a transport model (e.g. *MATsim*). Classical four-step transport models (trip generation, trip distribution, mode choice, and route assignment) relied on areal units as origin and destinations, designated as Traffic Analysis Zones or TAZ. Such models are, obviously, sensitive to the size and shape of the TAZ, has demonstrated by Chang et al. (2002); Zhang and Kukadia (2005); or Viegas et al. (2009). The current state-of-the-art, however, consists in activity-based models where individuals are the relevant level of analysis (Timmermans, 2003; Rasouli and Timmermans, 2013).

Such micro-simulation models are able to simulate individual decisions of travels (for multiple purposes) rather than only the number of trips between origins and destinations. Note that even if *MATsim*' internal principles are quite different from other transport models (Nagel et al., 2008), it is also an activity-based micro-simulation model. The coupling plug-in with *UrbanSim* is, however, limited to home-to-work commuting fluxes (see chapter 5). The integration between *UrbanSim* and *MATsim* (partially developed during the *SustanCity* project, see Nicolai and Nagel, 2015) appears, nevertheless, to be more complete and straightforward than for other transport models (e.g. METROSIM, see de Palma et al., 2015b).

The reader interested in details on transport modelling can refer to Axhausen and Gärling (1992), Bowman and Ben-Akiva (2000), de Dios OrtÃozar and Willumsen (2011), or Lucotte and Nguyen (2013). The point, here, is that such purely disaggregated transport model are nowadays in frequent use, including during the *SustainCity* project (see Nicolai and Nagel, 2015). On the contrary, the land use side of a LUTI model remains in most cases an aggregated one (see chapter 2). Even if individual agents are accounted for (as in *UrbanSim*), the space itself is still divided into areal units. As mentioned by section 1.1, the sensitivity of LUTI models to the size and shape of these areal units has never been assessed. Therefore, in the remainder of this thesis, the sensitivity analysis will be focused on this land use side of LUTI models.

1.3.2 Outline of the thesis

Part I: general introduction and state-of-the-art

This first part constitutes the theoretical background of the thesis. In the current chapter 1, we presented the motivation and aims of the thesis. Chapter 2 will provide a detailed overview of land-use and transport interaction models, by reviewing four successive topics: (1) the history of LUTI models' developments, (2) the internal principles of representative and/or currently operational LUTI models, (3) the spatial choices made in applications of these models for land-use or transportation planning, and (4) the theoretical background of spatial bias and their relevance for LUTI models.

Part II: sensitivity of LUTI models econometric' components to spatial bias

The aim of Part II is to improve the state-of-the-art on the sensitivity to spatial bias of econometric methods on which LUTI models rely to predict the evolution of a metropolitan area. It requires a sensitivity analysis of the *behaviour* of the model system to the *spatial extent* and *resolution*, in order to answer the following research question:

Research question 1 *Does a change in either the <u>spatial extent</u> or the <u>spatial resolution</u> influence the <u>behaviour</u> of a LUTI model? And, if so, by which mechanisms?*

LUTI models in general, and *UrbanSim* in particular, rely on two main econometric methods (see chapter 2): regression and discrete choice models.

The influence of the *scale effect* on parameter estimates of regression methods has been extensively studied, which is not the case for the *boundary effect*. For discrete choice models, the sensitivity analyses of the *scale effect* are much less extensive. Existing works do not focus on spatial bias, or use BSU larger than those on which recent LUTI models' applications rely (see section 2.5). Hence, the sensitivity analysis of the *behaviour* of *UrbanSim* to spatial bias will aim at advancing the state-of-the-art in the field of spatial bias in statistical methods, by considering questions that are relevant for LUTI models and have not (or poorly) been assessed in existing works.

In chapter 3, we study the sensitivity of an urban land price model, using a regression method, to the size and composition of the study area chosen to define the city. Chapter 4 assesses the sensitivity of DCM to the *scale effect*, with a particular focus on the implications in the context of LUTI models.

This part of the thesis uses the Brussels metropolitan area as an empirical case study. This city is a highly interesting case study to assess the influence of cities' delineations on econometric methods, due to its particular political context (see chapter 3). Nevertheless, prototypes *UrbanSim* applications developed for Brussels (i.e. Gallay, 2010; Patterson et al., 2010; and Patterson and Bierlaire, 2010) show that data availability is strongly limited for some crucial components of the model system (this issue is extensively discussed in Cabrita et al., 2015). Hence, the sensitivity analysis of *UrbanSim*' econometric components proposed in chapters 3 and 4 will be assessed using external data sets, rather than the database developed for the Brussels case study of *SustainCity*.

1.3.3 Part III: influence of spatial bias on LUTI models' outputs

LUTI models are, intrinsically, applied models. Part III focuses, therefore, on the the sensitivity of LUTI models' outputs to spatial bias. It attempts to explore both theoretical aspects and practical implications of the sensitivity of LUTI models to spatial biases, to answer the following research question:

Research question 2 *Does a change in either the spatial extent or the spatial resolution influence the outputs of a LUTI model? And, if so, can it jeopardise policy evaluation based on these outputs?*

The underlying assumption is that spatial bias will only constitute an issue for operational applications of LUTI models if the variations, induced by a change of either the *spatial extent* or the *spatial resolution*, are larger than (a) the inter-runs variations and (b) the variations due to the implementation of land-use or transport scenarios. Since this issue has received no attention in literature, until now, a complete analysis is required. It is the aim of chapter 5.

Development of an operational application of a LUTI model on a metropolitan area represents a tremendous amount of work (about 3.5 years in the case of the *SustainCity* project). Therefore, in chapter 5 a synthetic city is used to explore the sensitivity of *UrbanSim*' forecasts to both the *boundary* and *scale effect*. The influence of cities' delineation on LUTI models' outputs will, in particular, be assessed systematically. The practical advantages of this choice are that a synthetic case study allows defining BSUs of identical sizes (e.g. a grid) that can easily be re-aggregated into larger BSU levels. The influence of the size of the study area can also be examined more consistently, since the limit of the influence area of two CBDs can be fixed by the modeller (see chapter 5). The main interest, however, is theoretical. The synthetic case study allows defining the utility function of the agents. The set of feedbacks driving their location choices are, therefore, perfectly known, ensuring that variations in the outputs of the model observed between the different *spatial extent* or *resolution* are due to these spatial biases, rather than to noises resulting from the insufficient goodness-of-fit of the model.

The last step, conducted in chapter 6 aims to assess if the forecasted feasibility and/or sustainability of various land use or transport scenarios can be affected by the *spatial resolution* of the model (due to time constraints, this analysis has not been extended to the *spatial extent*). We rely on the Brussels case study of the *SustainCity* project for this latter analysis, since a real-world case study allowed for better estimates of the cost and revenues of the proposed scenarios.

1.3.4 Part IV: recommendations and conclusion

Part IV concludes the thesis. Hopefully, the analysis performed in Part II and III will provide two types of *practical insights* for modellers. First, the priority that should be devoted to the spatial bias issue, compared to other issues affecting the operational applications of LUTI models (see chapter 2). The second desired outcome is to be able to propose "best spatial practices", allowing reducing the sensitivity of LUTI model to spatial bias.

Therefore, chapter 7 will cover four successive aspects. (a) First, a summary of the findings of the thesis. Then, different recommendations allowing to reduce the sensitivity of LUTI models to spatial bias will be exposed, covering (b) the "best spatial practices" and (c) the potential developments of LUTI models' internal principles. Finally, (d) paths for further research are discussed, together with an alternative approach to LUTI models for land use and transport modelling. Figure 1.1 provides a graphical outline of the thesis.

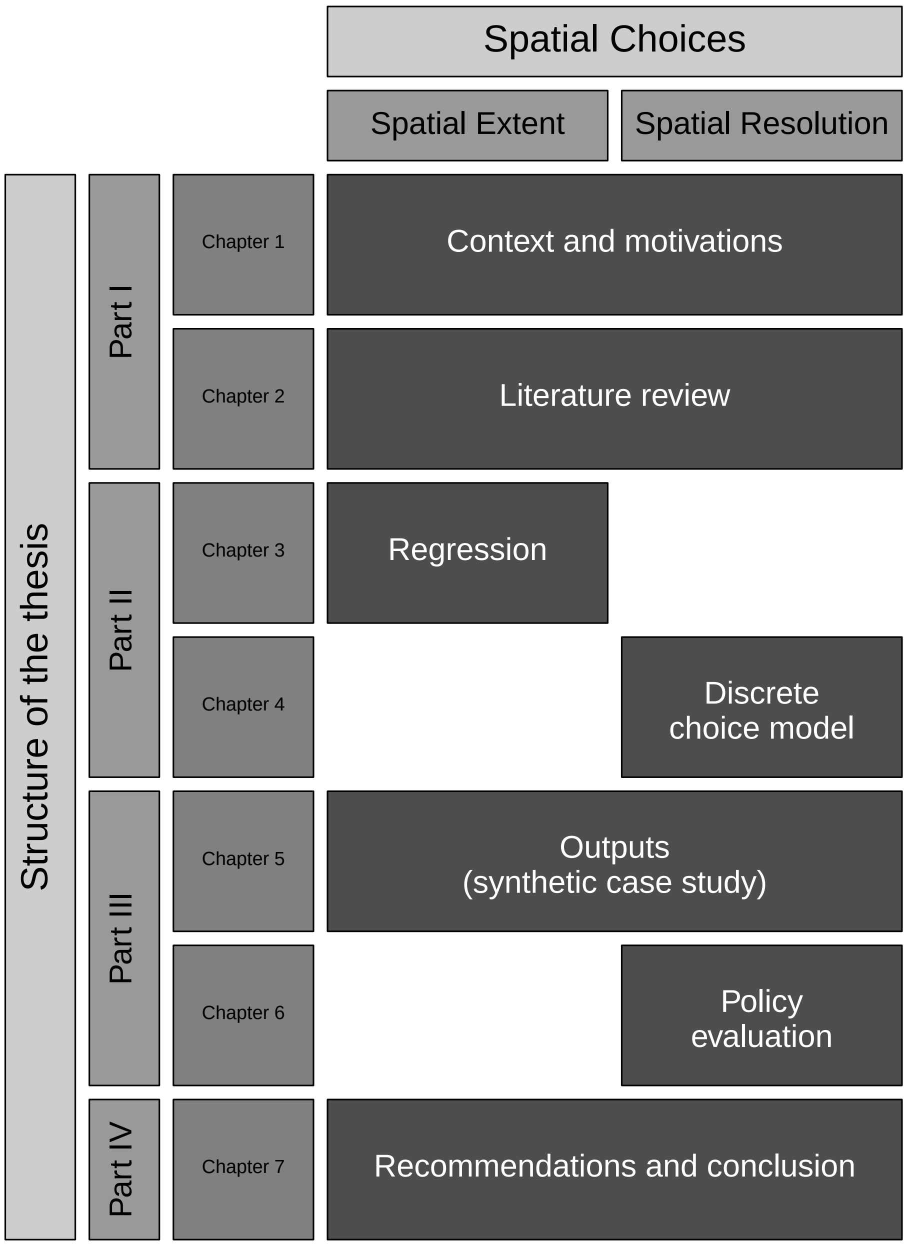

Figure 1.1 – Graphical outline of the thesis

CHAPTER 2

Space in Land Use and Transport Interactions models

2.1 Introduction

Even if the definition of Land Use and Transport Interactions models adopted in chapter 1 is restrictive, a large number of models fit into it. Their internal principles and development have been studied in numerous reviews (e.g. Wegener, 1995; Southworth, 1995; Miller et al., 1998; Miller et al., 1999; Wegener, 2004; Timmermans, 2003; Dowling, 2005; Hunt et al., 2005; and Wegener, 2014). The present chapter is largely based on these previous works. It intends, however, to propose an original perspective, by assessing how the space is taken into account within LUTI models. We focus first on the two main spatial choices that LUTI modellers face: (a) the delineation of the study area, hereafter referred as the *spatial extent* of the model, and (b) the choice of the Basic Spatial Units (BSU), i.e. the *spatial resolution*. Secondly, the influence of space on the internal principles of LUTI models will also be considered. Although the potential bias induced by the MAUP or the UGCoP are not new (see section 2.5), this chapter shows that they have been largely ignored in most applications of LUTI models (see also Thomas et al., 2015).

The chapter is divided into four successive sections. Section 2.2 presents an history of LUTI models, describing the early spatial challenges faced by these models and the current debates in that field. Section 2.3 introduces the internal principles of representative or state-of-the-art LUTI models. Section 2.4 proposes a meta-analysis of space within LUTI models applications based on two case studies: operational applications by Metropolitan Planning Organisations (MPO) in the US and applications of LUTI models in Europe published in peer-reviewed journals. Section 2.5 presents the conceptual framework of spatial bias and their relevance for LUTI models..

2.2 An history of space within LUTI models

The need to account for feedback effects in the land use and transport cycle was recognised in the early sixties (Hansen, 1959). But the history of LUTI models is made of cycles. Three waves can be identified (see Timmermans, 2003). The first operational developments (Lowry, 1964) have been criticised at the beginning of the seventies in the famous "requiem for large scale models"by Lee (1973). Various authors (Batty, 1994; Harris, 1994; Wegener, 1994) later argued that recent technological advances such as GIS had solved the seven sins outlined by Lee (1973). Lee (1994) rebuffed this statement, but LUTI models nevertheless prospered during the nineties. Note that the introduction of federal regulations in the US during the nineties, constraining Metropolitan Planning Organizations (MPO) to take into account land use effects in their long-term transportation plan (Waddell et al., 2007; Wegener, 2011a), also played a role. This was followed by a new cycle of criticisms, focused on conceptual limitations (Timmermans, 2003) or on the poor results produced by LUTI models (see e.g. Bartholomew, 2007; Wagner and Wegener, 2007; Nguyen-Luong, 2008). It does not seem that these criticisms lead to more usable models, but rather accelerated the development of dynamic micro-simulation models (Wegener, 2011a; Simmonds et al., 2013; Wegener, 2014). Nowadays, forty years after Lee (1973) seminal paper, the debate is still ongoing (see te Brömmelstroët et al., 2014; Batty, 2014): are we beyond the seven sins of large-scale models?

The three main steps that can be identified in the LUTI model's timeline are thus the following: (1) early development in the sixties, (2) renewed interest in the nineties, and (3) the recent introduction of micro-simulation models. The point of this section is not to propose a complete history of urban modelling during that period, for which we refer the reader to Harris (1985), Wegener and Furst (1999b), and Batty (2008). Instead, we propose to explore how the space was taken into account on each of these steps, on a conceptual point of view and for representative LUTI models.

2.2.1 From Lowry to Lee's requiem

The land use and transport feedback cycle

The underlying motivations for the development of LUTI models are classically described by the *land use transport feedback cycle* (Figure 2.1) that highlights the set of relationship between the location choices of agents and their travel behaviours. It originates from Hansen (1959). Space is a key aspect of this cycle, since transport is seen as the result of the spatial mismatch between residential location and economic activities (and conversely influence the location decisions of agents through transport costs). Although the purpose of this diagram is to provide a synthetic view of a complex phenomenon, many dimensions of space are over simplified by this land-use and transport feedback cycle. For instance, travel times and costs derive from distance between locations that are a function of both the spatial extent and resolution of the model: it has been demonstrated that the size of the Traffic Analysis Zones (or TAZ) influences the trip lengths in a travel model (see e.g. Zhang and Kukadia, 2005; Viegas et al., 2009). Location decisions of investors and users will also be constrained by several spatial characteristics other than travel times, such as land-use regulations.

Therefore, this example illustrates the simplistic representation of space in LUTI models frameworks. As we attempt to show in this chapter, the lack of interest for spatial biases can be observed in other theoretical grounds of LUTI models (e.g. the urban sub systems by speed of change, see section 2.2.2). In operational LUTI models, the *spatial resolution* is sometimes discussed, but mostly from a data availability point of view (section 2.4). The *spatial extent* almost never.

The Lowry model

The model of the Pittsburgh region developed by Lowry (1964) is often referred as the first land use and transportation model (Rodrigue et al., 2009; Wegener, 2014). This model represents the culmination of the idea that computer model were (at the time) new tools for planning (Harris, 1965) and constitute a breakthrough for urban planning (Wegener, 1994). The general principles of the Lowry's model still shape some of the current operational LUTI models and will be described in section 2.2. Let us simply recall that a gravity-type spatial interaction model is used to determine the equilibrium location of retail and residential activities (Rodrigue et al., 2009).

The spatial choices made by Lowry (1964) in the original application of his model to the Pittsburgh region are interesting. Due to data requirements, the model was fitted for the area covered by the Pittsburgh Area Transportation Survey of 1960 (PATS), which encompass (roughly) 1000 km^2, 1.5 million inhabitants and 550 000 jobs (Figure 2.2). It should be noted that the boundaries

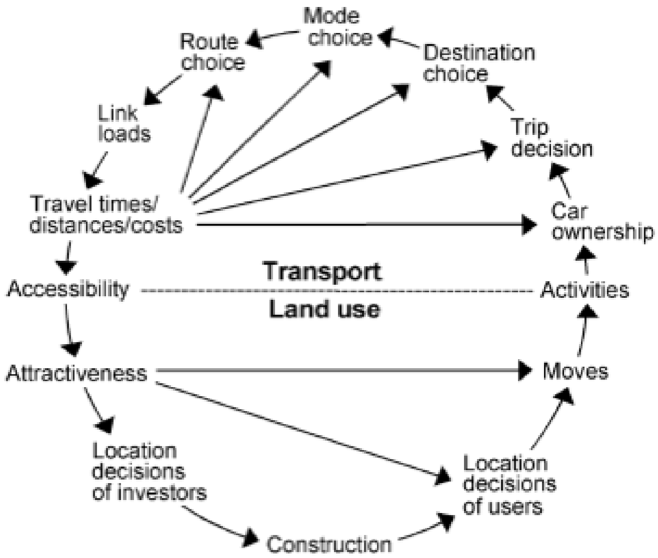

Figure 2.1 – The land-use and transport feedback cycle (source: Wegener and Furst, 1999b)

of the PATS, a cordon line of unclear nature, do not match exactly those of the Pittsburgh Urbanized Area defined by the 1960 census. This issue is mentioned by Lowry (1964) himself. Hence, the spatial extent of the project can be defined as an urban region based on a transportation criterion.

As Basic Spatial Units, Lowry (1964) uses custom grid cells: 456 units created by aggregating the PATS tracks on a grid with a one-mile interval (Figure 2.2). The reasons are that basic units from the PATS were too numerous and that Traffic Analysis Zone (also from the PATS) were overly specialized towards transportation study. The advantages and drawbacks of the retained zoning system are worth quoting: *"while the actual tracts boundaries follow those of constituent city blocks, the abstract grid takes no account of the boundaries of other natural areas; individual tracts may be functionally heterogeneous, divided by topography or discontinuities in land use. The main advantage of this geographic coding system is computational flexibility* [and] *is neutral with respect to theoretical patterns of urban structure"* (Lowry, 1964; pp. 58). The spatial resolution have thus been constrained by the modelling strategy rather than by data availability. Both the *spatial extent* and the *spatial resolution* can be criticised on a scientific point of view. It should however be stressed that such an accurate description of the spatial choices made during the development of

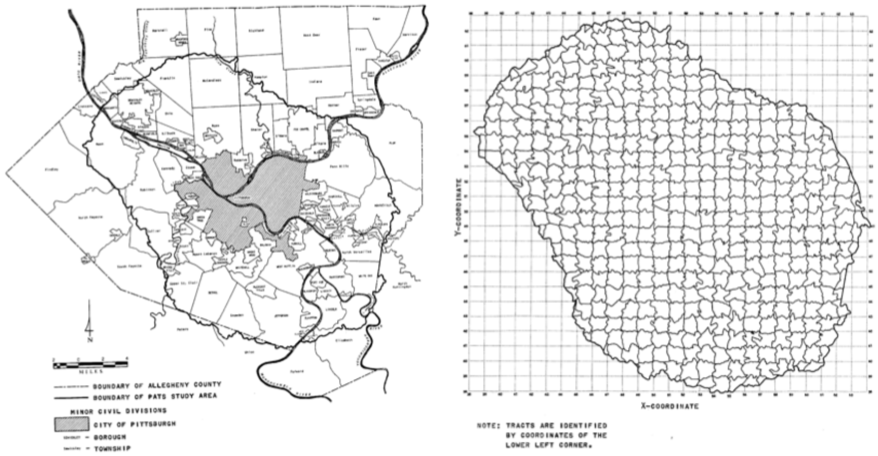

Figure 2.2 – **The Lowry's model of Pittsburgh** (left = study area; right = BSU; source: Lowry, 1964)

the model is rarely found in recent applications of LUTI models (see section 2.4).

Criticisms

Alongside with other urban models developed in the sixties, the Lowry's model has soon been criticized. Lee (1973) famous *requiem for large-scale models* effectively puts a stop to the development of land-use and transport interactions model for 15 to 20 years. It is worth recalling the "seven sins" identified by Lee (1973):

- **Hypercomprehensiveness:** large-scale models attempt to replicate too many processes and are expected to serve too many different purposes;
- **Grossness:** although a large amount of data is required, only aggregated results are produced;
- **Hungriness:** too much data are needed;
- **Wrongheadedness:** there is a hiatus between the processes supposed to be represented within the model and the actual set of equations driving the model;
- **Complicatedness:** the increasing number of variables leads to increasing interactions (feedbacks), many of them non significant, obscuring the behaviour of the model;

- **Mechanicalness:** numerical error may be present in computer's calculations;

- **Expensiveness:** large-scale models have very high costs.

A clear spatial aspect appears in three of them. (1) The grossness issue is largely a spatial one. It points out that large-scale urban models only produce aggregated results. According to Lee (1973), the usefulness of such results (e.g. the future number of inhabitants in a given zone) is limited to a small number of people involved in urban planning, and these people would not have needed the model to come up with similar forecast. To produce locally detailed results, a high spatial resolution is required but small and, therefore, numerous, BSUs increase the computational power needed. (2) Hungriness refers more specifically to this large amount of data required by LUTI models (a problem already recognized by Lowry, 1964). Spatial data add a further level of complexity in the data collection and processing task since they are often only available at different scales and for non-perfectly overlapping areal units (see section 2.5). Finally, (3) wrongheadedness states that the processes supposed to be represented by the model do not correspond to the equations or set of rules that are actually driving the model. One of the examples given by Lee (1973) is explicitly spatial: *"while valid at the scale of a metropolitan area the gravity model* [for trip distribution] *has no explanatory power at the neighborhood level"* (Lee, 1973, pp. 165) and results in the well-known ecological fallacy problem (see Robinson, 1950). According to te Brömmelstroët et al. (2014), most of these seven sins remain unsolved today, even if the development of GIS has somewhat simplified the task of urban modellers.

2.2.2 The Nineties, a golden age for LUTI models

In 1990, the US Congress adopted major amendments to the Clean-Air Act of 1970 (CAAA). The Intermodal Transportation Efficiency Act (ISTEA, 1991) soon followed this new environmental legislation. Both laws state that cities applying for federal funds for transport infrastructure must take into account the impact on land use and environment (Waddell et al., 2007; Wegener, 2011a). Note that they have been preceded by other legislations with similar purposes, e.g. the Federal-Aid Highway Act (1962) and the National Environmental Policy Act of 1970 (Bartholomew, 2007), and followed by others (e.g. TEA-21 of 1998; SAFETA-LU of 2005). Nevertheless, CAAA and ISTEA are frequently presented as the cause of the renewed interest for LUTI models in the nineties (see e.g. Dowling, 2005; Waddell, 2011).

More technical reasons can also be highlighted. Computers made large progress between the Lowry's model and the nineties. The most notable development, however, was the Geographical Information Systems, or GIS. These tools were soon used into LUTI models (see section 2.2.2) and to build various GIS-based environmental models (Goodchild et al., 1993; Goodchild and Marvin, 1996; Sui, 1998)[1].

The consequence has been the development of a large variety of models. Wegener (1994) found 13 different LUTI frameworks. This number later grows to 17 (Wegener and Furst, 1999b) and even 20 (Wegener, 2004). Note that these reviews focus on LUTI models with an academic origin, and do not include many custom frameworks developed by MPO (see section 2.4). Southworth (1995) identifies 17 LUTI frameworks, corresponding only partially to those found by Wegener (1994). By combining these reviews, a total of 30 different models emerge (Table A.1, in Appendix). Most of them, however, have not reached an operational status, or are not anymore in use today. Therefore, section 2.3 will only provide an overview of the internal principles for those that are still in use nowadays (see also Table A.3, in Appendix).

Dynamics of urban sub systems

The model of Lowry (1964) was an equilibrium model: the state of the city is predicted for a one-time point in the future. The evolution between the base year and this horizon is not modelled.

However, modellers will soon advocate that processes affecting the urban system react differently over time to a perturbation, i.e. that "Rome was not built in a day". Table A.2 presents the conceptual formalisation of the urban sub systems by speed of change (Wegener et al., 1986; based on Snickars et al., 1983). Note that among the processes described by Table A.2, the slower ones are those with the largest impact on the land-use. *"Human settlements evolve over a long time span by the cumulative efforts of many generations. The resulting physical structure of cities displays a remarkable stability over time prevailing even after major devastations such as wars, earthquakes, or fires"* (Wegener et al., 1986; pp. 4). Hence, many LUTI models (see section 2.3) are now quasi-dynamic models and proceed by iterations (of usually 1 year), meaning that the final state in iteration t is the initial state in $t+1$ (Simmonds et al., 2013; Wegener, 2014).

Nevertheless, another group of modellers (mostly represented by A. Anas, see Anas and Liu, 2007) continued the development of equilibrium LUTI models, and the two approaches co exist nowadays. This thesis focus, however, on the "dynamic" family, represented here by *UrbanSim* (see chapter 1).

[1]Note that the development of cellular automata models of land-use (e.g. White and Engelen, 1993; Batty, 2005) can also be related to GIS.

The California Urban Futures model

This model succeeds to the Bay Area Simulation Systems (itself descending from the Lowry's model, see Goldner, 1971) and was continuously updated during the nineties (Landis, 1994a; Landis, 1994b; Landis and Zhang, 1998a; Landis and Zhang, 1998b). It provides a good example of the evolution followed by LUTI models throughout that period. Note that it presents several differences with the models originating from Lowry (1964) framework: (1) the population growth is allocated to discrete individual sites rather than to large areal units. (2) Accessibility to jobs is not the only factor affecting the location and density of new real estate development. (3) The model takes advantages of GIS techniques to organize a spatial database and manage land development potential. (4) Development policies are explicitly implemented in the model. Finally, (5) the model is easy to use and includes a graphical output. Moreover, the model is quasi-dynamic and simulates the future land-use by time-steps of five years (Landis, 1994a).

The model has been developed for the Northern California Bay Region (19 counties surrounding the San Francisco's Bay). Landis (1994a) insists on the need for the model behaviour and output to be relevant for planners and policy makers. Since policies undertaken by various public authorities had to be simulated, the model had to use administrative delineations. Three levels of basic units are used: counties, cities, and Developable Land Units (or DLU). Counties are the largest areal units of the model and correspond to administrative units. They are divided among cities (i.e. a municipality) and non-incorporated areas (i.e. land that is not organised as a self-governing entity). The DLUs are custom units. They *"do not have regular shapes or sizes, but are generated as the geometric union* [in a GIS] *of different map features and their attributes"* (Landis, 1994a; pp. 408). The layers used to define the DLUs include administrative boundaries, cities' sphere of influence, wetland, slope, current land-use types, road network, and land prices (Landis, 1994a).

Figure 2.3 shows the principles of the model. An Ordinary Least Squares regression predicts the future population per counties and cities. Independent variables include the population in $t-5$ and different place specific factors. The population growth of non-incorporated areas is the difference between the future population per county and the total population of the cities in that county. The spatial allocation sub model then develops the DLUs (i.e. simulate the construction of new residential units), starting from the more profitable one, until the total future population is allocated. The size of the DLU, its zoning rules if any, and the population density in similar, already developed, areas determine the number of inhabitants per DLU. The profitability of the DLUs is estimated as the (future) houses selling price minus all development-related costs: land price, construction, public infrastructure, etc (Landis, 1994a). This framework allows an easy implementation of new policies, by adding into the spatial database a layer (e.g. greenbelt, land-use rules) accounting for this policy (see Landis, 1995 for examples).

Landis (1994a) notes, however, that several features of the model could be improved, such as more comprehensive development rules, the incorporation of employment, and the distinction between different types of households. A second version of the model, known as CUF-II, has been developed to remedy to these shortcomings (see Landis and Zhang, 1998a; Landis and Zhang, 1998b). In this latter version, the growth sub model is divided between two categories of households and 10 employment sectors. DLUs have been simplified to grid cells of 100 by 100 meters. Landis and Zhang (1998a) note that parcels would have been a near-ideal unit to study land-use changes but that such data were not available, hence the grid cells. Moreover, rather than deterministic rules, CUF-II implements a multinomial logit framework to predict land-use changes. Nine different development events (e.g. undeveloped to single-family residential use) are defined. The utility function depends on the initial land-use, grid cells characteristics (slope, accessibility to other activities), local zoning rules, etc. Transition probabilities estimated by the logit model are used as bids, representing competition between land-uses in a given site (Landis and Zhang, 1998a).

2.2.3 21^{st} century: the age of reason?

The trends towards agent-based micro simulation models exist since (at least) the second half of the nineties. Note that most LUTI models used nowadays were already operational at the time (see section 2.3). It is clear, however, that the enthusiasm of the Nineties has been tempered since the new millenium, even if the need to account for land-use/transportation interactions in urban planning is recognised (Badoe and Miller, 2000).

Two reasons support these recent criticisms. The behavioural and theoretical grounds of LUTI models are still weak (Timmermans, 2003). Land-use and transportation field borrows methods and tools from other disciplines and put them into one integrated model. As a result, *"it is a strange experience to notice that at symposia on integrated land-use - transport systems often basic principles that were discussed (...) considerable time ago are still high on the agenda"* (Timmermans, 2003; pp. 21). Moreover, different large-scale projects based on LUTI models have failed to meet their goals (e.g. Wagner and Wegener, 2007, Nguyen-Luong, 2008, or Wegener, 2011a). In particular, the time, cost, and complexity of building the spatial database needed by the model is often under estimated.

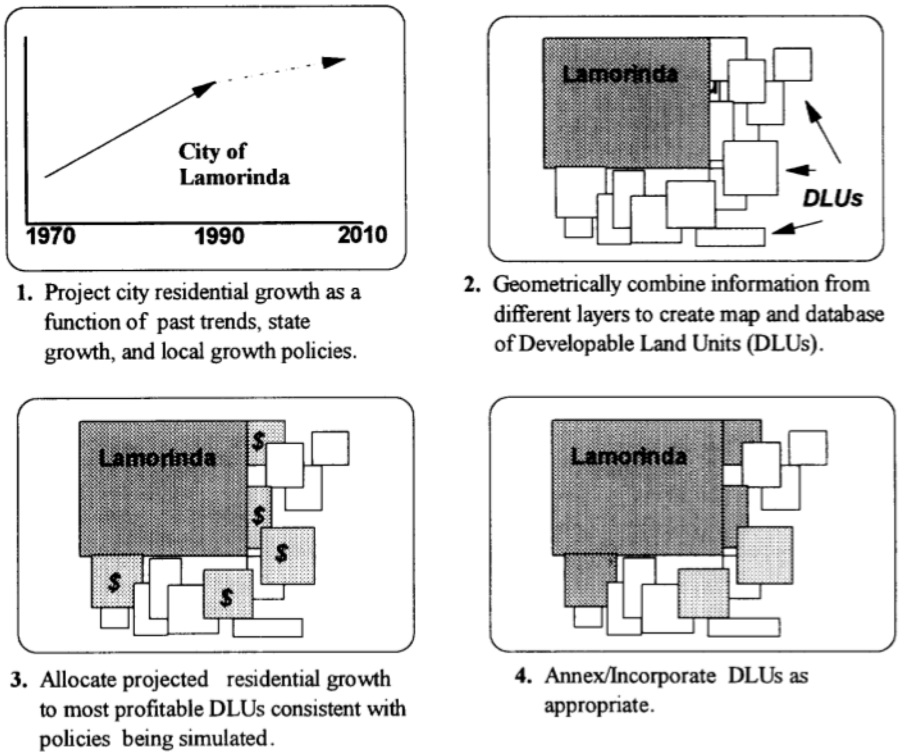

Figure 2.3 – The California Urban Futures model (workflow from Landis, 1995)

Current practices

A review of land-use and transport scenarios planning from Bartholomew (2007) is presented in section 2.4. All projects reviewed in this paper are located in the US and took place between 1989 and 2003. The modelling framework is generally less integrated than academic models reviewed by Wegener (1994) or Wegener and Furst (1999b). In most cases (Table 2.4), it consists in a travel model to which a GIS land-use sub model has been added (in a similar fashion than in the CUF model). Sensible differences are visible in today's modelling practices. Lee (2009) conducted a survey focused on TMA[2] and MPO responsible for an area having more than 200 000 inhabitants (total of 201 organizations, among which 146 responded). The results show that 47% the TMA/MPO where doing both land-use and transport modelling (while 46% only made transport modelling). MPOs commonly use models that originate

[2] Transportation Metropolitan Agency

Land use model used	%
Home-grown model	23
UrbanSim	15
GIS-based model (e.g. UPLAN)	12
PECAS	9
Qualitative policy judgement	9
DRAM/EMPAL	6
Other model	16
No data	10

Table 2.1 – **Land-use modelling tools of large TMA/MPO** (source: Lee, 2009; n = 68)

from academic (DRAM/EMPAL, UPlan, UrbanSim, PECAS), as showed by Table 2.1. The transport model used was for 73% of the 146 MPOs a classical four-step model, only 4% of the planning organizations using an activity-based model. Moreover, despite regulatory constraints, merely 27% of the MPOs achieved an operational integration between land-use and transport (and 12% were working on it). It should also be mentioned that Lee (2009) found that large MPOs were more likely to use advanced quantitative models than small one.

Current debates

Wegener (2014) identifies two debates, or dividing lines, in the current urban modelling frameworks: (1) the opposition between equilibrium and dynamic models, and (2) the aggregate macro-economic approach versus disaggregated models with an individual representation of agents. To our point of view, the first debate has largely been solved in favour of (quasi-)dynamic models, even if no truly dynamic (i.e. having a continuous representation of time) urban models exist. The second debate is more problematic and has a stronger spatial component.

It results from the influence of the micro simulation approach (see Orcutt et al., 1961) on LUTI models. Contrary to various transport models (including *MATsim*, see Nagel et al., 2008), land use models such as *UrbanSim* are not a real micro simulation model. Nevertheless, modellers have internalised the idea that individual units (households or jobs) should be accounted for, leading to the development of disaggregated urban models with a distinctive representation of agents. This evolution was also made possible by the growth of computer power and the increasing availability of detailed spatial data in the Nineties (Wegener, 2014).

This approach is grounded in micro-economic theories. It also allows taking into account the heterogeneity between agents and their preferences. Under that conceptual framework, the use of discrete choice models to forecast location choices of agents or of new real estate development is a natural choice (Orcutt et al., 1961; Heppenstall and Smith, 2014). However, disaggregated models results *"are subject to stochastic variations, i.e. may differ significantly between model runs with different random number seeds unless averaged to a level of aggregation they were designed to overcome"* (Wegener, 2011a; pp. 2). Stochastic variations (or Monte Carlo error) may be greater than the variations observed between scenarios, precluding the use of the models as planning tools. Their magnitude is a function of the number of choices (e.g. the number of households moving) and the number of alternatives (e.g. the number of zones) simulated. Wegener (2011a) shows that stochastic variations increase when the ratio choices/alternatives decrease. For a fixed number of choices (i.e. if we assume a constant population), the ratio choices/alternatives will decrease if the number of alternative increases. Hence, for a given city, small BSU leads to larger stochastic variations than aggregated ones.

The presence of stochastic variations has been demonstrated in operational LUTI models, in particular *SimDELTA* (Feldman et al., 2007) and *UrbanSim* (Sevcikova et al., 2007), and in other applications of discrete choice models (Moeckel, 2007). Wegener (2011a) outlines four potential solutions to this issue: (1) aggregate the output at a larger level than the BSU used by the model. (2) Artificially increase the number of choices (see Hunt et al., 2008). The most often recommended solution, and the one applied in this thesis, is (3) to run the model several times and to take the average results. Finally (4), the stochastic nature of micro simulation models can be recognized, and results presented as probabilities of transitions rather than deterministic forecasts.

2.2.4 The future?

The history of urban modelling goes from macroscopic equilibrium model to agent-based micro-simulation models (Batty, 2008). Today's models represent more detailed process, at a finer geographical scale. However, methodological constraints of disaggregated models are numerous (see Nguyen-Luong, 2008; Wegener, 2011a): data requirements, computing time, and stochastic variations. Since these challenges are case-specific, it is unlikely that a general solution will emerge in a near future. Potential future development of LUTI models, and their adequacy to solve spatial bias, are discussed in chapter 7.

2.3 Representation of space within LUTI models

Over time, LUTI models have known several evolutions of their internal principles. A clear trend can be identified, from older aggregated equilibrium models towards newer disaggregated dynamic models (section 2.2). Hence, the aim of this section is to assess in more details how the space is represented in the internal principles of LUTI models. Describing every existing LUTI (for a non exhaustive list, see Table A.1) is obviously not possible, since many of them are tailored developments for one specific case study. Moreover, very few of them have reached the state of *operational* (i.e. having at least one practical application), *comprehensive* (which includes most spatial processes, i.e. land development, location choices of agents, and transport), *integrated* land-use and transport frameworks (Hunt et al., 2005). Table A.1 shows that the number of LUTI models compared in recent reviews (Hunt et al., 2005; Simmonds et al., 2013; Wegener, 2014) is much more limited than those listed by Wegener, 2004. The number of models that have been applied on several case studies is also highly limited (note that most of them, including DELTA, PECAS, TRANUS and UrbanSim, have in common to be supported by a commercial consultancy company).

Therefore, we rely on two criteria to select the LUTI models that will be reviewed here. First, they have to constitute a representative sample of the different modeling strategies implemented. Table A.3 presents a short summary of the frameworks used in the models reviewed by Wegener (1994). In the same way, Table 2.2 proposes a typology of LUTI models based on Timmermans (2003) and Wegener (2014). Note that the third category of Timmermans (2003) only differs from the second by relying on a transport model simulating activities rather than solely home-to-work trips. Secondly, reliable description of these models should exist in the literature.

The eight models selected here are, by chronological order, the Lowry model, MEPLAN, IRPUD, TRANUS, DELTA, MUSSA, UrbanSim and PECAS. Their internal principles are presented in appendix A, on the exception of UrbanSim. This latter model is used in the analyses (see chapter 1) and it will, therefore, be extensively presented in chapter 5. Note that this selection left aside various models tailored for one specific case study. Note that relying on the international literature may have bias this selection towards model originating from Anglo-Saxon academics and/or consultancy company (excluding Europeans developments such as PIRANDELLO).

2.3.1 Overview of the representation of space

All models reviewed here operate on "zones" (defined as non overlapping basic spatial units of irregular size and shape), rather than on pixels or custom GIS-based overlay (as in, for instance, the CUF model). Table 2.3 summarises their nature. In terms of size, three groups can be drawn. MEPLAN and TRANUS both rely on a small number (50 to 100+) of large land use zone (Echenique, 2001; Hunt et al., 2005). The Lowry (1964) model (see section 2.2.1), MUSSA, (Martinez and Donoso, 2010) and UrbanSim use relatively small BSU (see

	Wegener (2014)	
Timmermans (2003)	Spatial interaction location models	Accessibility based location models
Aggregated spatial interaction models	Lowry, DRAM - EMPAL	IRPUD
Utility maximising multinomial logic based models	MEPLAN, TRANUS	DELTA, MUSSA, *UrbanSim*, PECAS
Activity based micro simulation models		ILUTE, ILUMASS, RAMBLAS

Table 2.2 – Typology of LUTI models (based on Timmermans, 2003 and Wegener, 2014)

chapter 5). The latter appears to be the more spatially disaggregated model among those reviewed here. Finally, IRPUD (Moeckel, 2007; Wegener, 2011b), DELTA (Simmonds and Feldman, 2005; Bosredon et al., 2009), and PECAS (Hunt et al., 2009a and Hunt et al., 2009b) implement a multi-level representation of space, with large zones further divided into small one. For instance, there are three nested levels in IRPUD: (1) a macroscopic model of economic and demographic changes, (2) a mesoscopic model of households, jobs, and real estate development location choices, and (3) a microscopic model of land-use changes (Timmermans, 2003).

In terms of nature, most models rely on administrative units (Table 2.3). Custom units are generally re aggregation of census track and are, therefore, also based on administrative units. Note that the Traffic Analysis Zones may differ from the land use zones. Overall, no constraints are given on the size, shape, or nature of the BSU except some technical one such as a limited number to reduce computation time.

2.3.2 A typology of LUTI models

A clear trend appears in LUTI model, from macroscopic to microscopic model (section 2.2). It can be decomposed in two components: the evolution towards (1) more detailed representation of agent's behaviour, and (2) smaller basic areal units. Figure 2.4 proposes a qualitative typology of the LUTI model reviewed in section 2.3 based on these two disaggregation components. The first one relates to the inclusion into LUTI models of more detailed representation of the agent's behaviour. Examples include the housing market sub model of IRPUD and the space development sub model in PECAS.

Model	BSU Size	Type	**Example**
MEPLAN	Large	Administrative	Bilbao: 66 BSUs for 1 500 km^2
TRANUS	Large	Administrative	50 to 100+
Lowry	Small	Custom	Pittsburgh: 456 BSUs (aggregation of survey track)
MUSSA	Small	Custom	Santiago de Chile: 409 BSUs
UrbanSim	Small	Administrative	Paris: 1 281 BSUs, μ = 9.3 km^2; Brussels: 4 945 BSUs, μ = 1.04 km^2
DELTA	Multi-level	Custom	Scotland: 50 "areas" divided into 720 "zones"
IRPUD	Multi-level	Administrative	Dortmund: 246 BSUs + 54 external zones
PECAS	Multi-level	Administrative	Maximum 750

Table 2.3 – Typology of BSU in LUTI models

The second one, spatial disaggregation means the evolution towards small basic spatial units, which culminates in UrbanSim. It seems to have followed the increasing availability of detailed spatial data (thanks to GIS technologies) and of computing capacity. Hence, the spatial resolution of LUTI models appears to be mostly constrained by data availability even if data requirements of LUTI models are often poorly described. Although this spatial disaggregation trends is valid on the long term, a distinction should be made between models that (a) focus on small basic spatial units (*UrbanSim*) and (b) those who rely on a multi-level representation of space (IRPUD, DELTA, PECAS).

The main econometric methods implemented in LUTI models are (a) spatial interaction models, (b) discrete choice model, and (c) linear regression model (see appendix A). Section 2.5 describes the conceptual framework of the sensitivity of these methods to spatial biases. Our point, here, is that these internal principles are not independent of the level of spatial aggregation used by LUTI models. On the first hand, spatial input-output matrix limits the level of spatial details since the coefficient of the matrix become unreliable for small zones (Hunt et al., 2005). On the other hand, Discrete Choice Model (DCM) and regression require a sufficient number of observations to produce robust

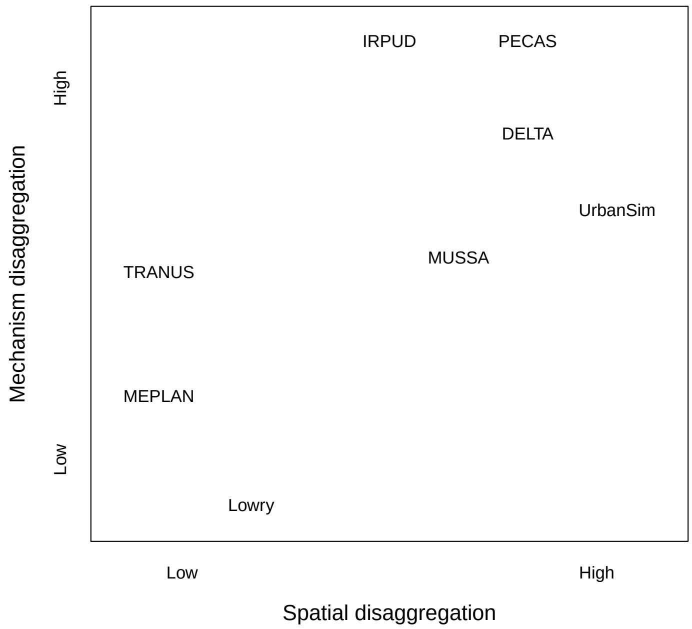

Figure 2.4 – Typology of LUTI models (horizontal = spatial details; vertical = internal principles details)

results. Therefore, the risk exist of using a model whose internal principles are inconsistent with the level of spatial aggregation of the case study.

Overall, these short descriptions of a representative set of LUTI models show that space is never seen as an issue. No limitations or guidelines for the choice of the BSU are given in the documentation of LUTI models, except for purely technical constraints (computation time). All models are presented as flexible and it is claimed that they can be adapted to every context. The fact that most of these models are supported by private consultancy company, and constitute therefore a commercial product, certainly influence this vision. Note that Timmermans (2003), Dowling (2005), Hunt et al. (2005), or Wegener (2014) have proposed more detailed comparisons of LUTI model based upon their level of time, spatial and economic details.

2.4 Spatial extent and resolution: a meta-analysis

Section 2.2 and 2.3 show that even if LUTI models are, by nature, spatial, the zoning system used does not appear to be a key component in the development of these models. Nevertheless, spatial issues raise specific challenges for the data collection and processing, as well as for the comparability of the results among case studies (Thomas et al., 2015). To characterize the representation of space in operational applications of LUTI models, this section attempts to identify their spatial extent (i.e. the size and composition of the study area) and resolution (i.e. the size of the BSU). An exploratory analysis is proposed for two case studies: (a) the use of LUTI models by Metropolitan Planning Organizations (MPO) in the US and (b) applications of LUTI models in Europe published in peer-reviewed journals.

2.4.1 Land use and transportation planning in the US

Data collection

Since 1962, all urbanized areas of more than 50 000 inhabitants are required to have their own Metropolitan Planning Organisation, or MPO (23 U.S. Code Art. 134 - 135). They are public agencies, composed of representatives of local governments and transportation authorities responsible for land-use planning and long-term transport management. Land use and transportation planning from MPO are almost exclusively published as grey literature (i.e. technical reports or summary for policy makers) which raises several difficulties. The first one is that such works are not indexed on scientific bibliographic databases (e.g. ScienceDirect, Scopus). Hence, we use here a sample of 79 projects of land-use and transport planning (for a total of 62 metropolitan areas, see Figure 2.5) collected by Bartholomew (2007). It covers a wide range of applications, from re-development of urban brown field to long-term planning for large metropolitan area. The sample was collected by mean of a two-stage survey conducted in 2003 - 2004 (see Bartholomew, 2007 for details). The key point is that it was directed towards organization's members of either the National Association of Regional Council or of the Association of Metropolitan Planning Organizations, to which the author asked if they were conducting scenarios planning projects in the field of land use and transport.

The aim of the work of Bartholomew (2007) was to study scenario planning, not to review spatial choices made by modellers, which ensure the independence of the sample. Three drawbacks should, however, be mentioned regarding the use of this data set for the goal pursued here. First, a clear selection bias exists, since it only includes projects initiated by planning organization, and not those that originate from academic. Secondly, most of these projects were conducted in the Nineties (50% of them being completed before 2001 and 75% before 2003), excluding recent developments. Finally, most of the projects in

Type of tools	**n**	**%**
Travel forecasting model	47	55
- with transit/pedestrian-oriented development sub model	9	11
- with a GIS scenario building tool	20	25
- with a land use allocation model	7	9
Sketch travel model	3	3
Sketch land use/travel model	3	3
Land use model only	4	5
GIS model only	10	12
Economic model/analysis	6	7
Other and no data	13	15

Table 2.4 – Analysis tools used in scenario planning projects (table from Bartholomew, 2007)

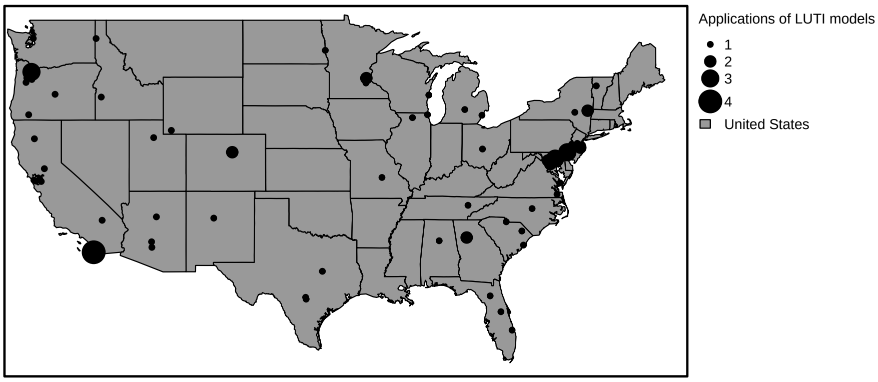

Figure 2.5 – Land use and transport scenario planning projects (sample from Bartholomew, 2007; the size of the dots is function of the number of projects)

the sample focus on transportation scenarios and, as a consequence, the tools used are often not true LUTI models (see Table 2.4). Nevertheless, Figure 2.5, showing the spatial distribution of the projects reviewed by Bartholomew (2007), is consistent with the distribution of the largest cities in the US and with the population density. The over-representation of Oregon can be explained since it is the state of origin of the author. Hence, the sample does not appear to suffer from a geographical selection bias.

Methodology

Among the sample collected by Bartholomew (2007), we have been able to retrieve the original references for 48 of the projects: 38 technical reports, nine summaries for policy makers and one paper in a peer-reviewed journal. Moreover, a technical appendix of the work of Bartholomew (2007) is available online and describes each project (see Bartholomew, 2005). From these documents, different information have been extracted (see Table 2.5) to assess the spatial extent and resolution of each project.

The main difficulty of technical report and summary for policy makers is that their contents vary widely from one document to another. On average, the level of details decreases from a technical report to policy-maker brief and then to the annotated bibliography provided in Bartholomew (2005). The accuracy of the data also differs from one variable to another. As a result, eight projects have been excluded due to a lack of data (none of the variable given in Table 2.5 could be gathered from their reference). It should also be noted that the 71 remaining projects (see Table A.4) corresponds to only 56 different case studies (i.e. study area). In particular, there are four projects in San Diego and three for Baltimore, the Willamette Basin and Wilmington (Figure 2.5).

Even for the remaining projects, missing values are frequent (Table 2.5). Variables related to the spatial extent are complete for 48 projects and have a fairly good level of accuracy, on the exception of population data. The numbers given for this variable were those found in the reference of the project when possible and if not those of the 2000 census (since the starting point of the project is in most cases not given). Hence, these values should be considered as an order of magnitude, for comparison purpose, rather than accurate figures. The accuracy of variables related to spatial resolution (results and BSU) is on average lower. Only 23 projects have complete information, while the same number has missing value for the two variables. In a given work, results are often presented at different levels (e.g. at the study area level for transport indicators but at the county level for demographic or land-use projections). Moreover, the BSUs of the model are generally poorly described, even for simple characteristics such as their total number or their average area. They should hence be seen as qualitative indicators. Overall, data for all of the variables given in Table 2.5 have only been obtained for 17 of the projects, and 24 projects have one missing value.

Spatial extent

Six categories of projects can be identified based on the composition of the study area and the criteria used for its boundaries (Table 2.6). Administrative and MPA categories are geographically similar, since their boundaries often follow those of the counties. The difference is that MPOs are constrained to use the metropolitan planning area boundaries as study area, while for other

Field	**Definition**	**Data**
Study area	Name of the study area	71 (100%)
Composition	Qualitative description of the the study area, i.e. main cities or number of counties	71 (100%)
Methodology	Criteria used to define the study area	56 (79%)
Surface	Total surface (in km^2) of the study area	54 (76%)
Population	Number of inhabitants in t_0	65 (91%)
Results	Level of aggregation used in to present the results	36 (51%)
BSU	Type and number of basic spatial units	35 (49%)

Table 2.5 – Spatial characteristics of LUTI scenario planning projects (sample from Bartholomew, 2007; Data = number of projects for which the information could be gathered)

organizations the use of administrative boundaries is a matter of choice. Only one project (Albuquerque) uses a delineation of the metropolitan area based on a functional indicator, corresponding to the extension of the water distribution network (leading to a morphological area). The reason given by the author is that the city has sprawl out of it's administrative boundaries and that the water service area corresponds to the zone covered by urban services. There is a clear relationship between the criterion used to define the study area and its composition. The Development criterion designates both the extent of the project and the nature of the area of interest. In a similar way, for all projects in the Network category, the study area can be qualified as a corridor, i.e. an area that has no other intrinsic characteristics than to be at less than a given distance from the studied transport infrastructure[3]. The study area of the Albuquerque project, only member of the Functional category, corresponds to a metropolitan area. For most projects, however, administrative boundaries are used. In the MPA group, 20 of the projects' area of interest encompasses several counties (between 2 and 11), while for the city of Flagstaff it consists in its metropolitan area (note that this relatively small city is surrounded by the second largest county of the US, Alaska excepted). The use of counties boundaries by MPO is not mandatory, but it seems to be always the case for our sample. A larger diversity appears for the Administrative group, with three projects corresponding to cities or township, five to one county, 14 to several counties, and two to the state of New Jersey. Hence, this latter category do not always correspond to metropolitan area but also include regional applications.

[3]The term "corridor" is used in different references to designate the project, e.g. Highway 41 corridor master plan or Mountain View corridor growth choices study.

The projects vary widely in scale, with a study area ranging from 0.08 to 165 000 km^2 and a population varying from 0 (development projects) to about 18 millions. The population density (for the 50 projects for which both the surface of the study area and the population are known, see Table A.4) is comprised between 0 and 1 385 inhabitants per km^2. Note that the projects' average population density (268 hab/km^2) is significantly lower (t = -2.91**) than the threshold used by the US Bureau of Census to define urban areas (390 hab/km^2). It suggests that most study area exceeds these urban areas and rather corresponds to the metropolitan statistical area (see Federal Register, 2000 for further details on the U.S. Census Bureau definition's of urban areas). It should also be noted that the distribution of both the surface, and the population, is highly skewed to the right (skewness of 3.8 and 2.7). This is due to the presence in the sample of two projects (Southern California and Chesapeake Bay) that correspond to region rather than to metropolitan area. Some relation can be highlighted between the scale of the projects and the definition of their study area (Figure 2.6). The Network category consists in all case in impact study of transport infrastructure improvements and, therefore, have a smaller scale than projects based on administrative or MPA delineations. Development and Functional categories had to be excluded, due to a lack of data, but also correspond to small-scale projects. Projects based on administrative or MPA delineations do not show significant differences in terms of scale.

To sum up, justifying the spatial extent of the project does not appear to be a necessity to modellers and/or consultant involved in its realization. It can be understood for Metropolitan Planning Areas, since their boundaries are defined by legislative prescription and *"shall encompass at least the existing urbanized areas and the contiguous area expected to become urbanized within a 20-year forecast period"* (23 US Code Art 134). The area for which MPOs are responsible is thus defined in their status, making the question of the spatial extent irrelevant. Nevertheless, it is not mandatory that the boundaries of this area follow those of the counties that are members of the MPO. The fact that this is always the case in our sample suggest that delineations of the MPOs' responsibility area is not based on an analysis of the extension of the metropolitan area, but rather on a default choice. In a similar fashion, projects belonging to the Administrative category are in most cases sponsored by a public administration. It is likely that this sponsor will require results for all the areas under its jurisdiction, and the degree of freedom left to modellers is unknown. For Development and Network projects, however, this lack of interest is much more problematic. They correspond to relatively small-scale projects, and one may thus wonder if the study area defined really encompasses the area that will be impacted by the project. In particular, corridor study areas are geometric delineations (e.g. all areas at less than three miles from the studied highway) and therefore assume an isotropic influence of the project. Finally, United States' counties vary widely in size over the country. The scale of the

Criterion	Definition	Projects
Administrative	Administrative delineation, usually counties	24 (34%)
MPA	Metropolitan planning area of the MPO sponsoring the project	21 (30%)
Network	Road or rail network	11 (15%)
Development	Area of urban development project (brown field valorisation or new construction)	5 (7%)
Watershed	Catchment area of a river	5 (7%)
Functional	Study area based on functional criteria	1 (1.5%)

Table 2.6 – Study area of the land use and transport planning projects (sample from Bartholomew, 2007; n = 56)

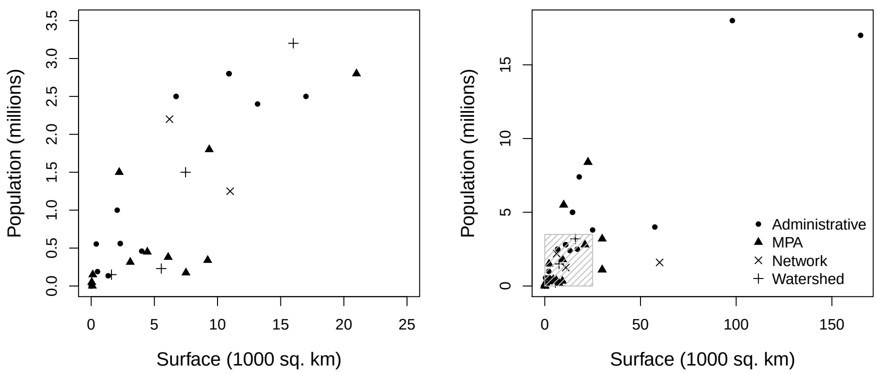

Figure 2.6 – Land use and transportation planning projects by study area and population (sample from Bartholomew, 2007; the grey rectangle denotes the area covered by the left plot)

project is not thus only function of those of the metropolitan area studied, but is also affected by the administrative delineations. The surface of the study area may thus differ widely for two cities having a similar population (see Figure 2.6). Hence, the nature of the study area is likely to vary, reducing the comparability of the results.

Spatial resolution

Due to the high frequency of missing data, the sample is here reduced to less than 30 observations. No statistical tests could be conducted and this latter step is thus only qualitative. For transport indicators (e.g. modal split, evolution of fluxes), results are mostly presented at an aggregated level, either the study area or per county, even when the model uses smaller traffic analysis zones. Land-use changes are generally given at an aggregate level (e.g. evolution of the built-up area per county). Maps at a smaller level can be found, but are used for illustration purpose only. The choice of the BSU is never discussed, nor than the BSU themselves. None of the references include simple indicators such as the average size of the BSU, and only four gave their total number. Moreover, the BSUs often vary between the land use and the transport component of the modelling framework. Three types of BSU can be identified: (1) census blocks, (2) Traffic Analysis Zones (TAZ) and (3) GIS overlay. Census block are the lowest geographic unit defined by the US Census Bureau and, consequently, the smallest unit for which statistical information's are available. TAZ are custom geographic units, defined by modellers to run a travel model. They usually correspond to aggregation of census blocks. Finally, land use is often available as raster data. As a result, different projects uses GIS overlay to merge statistical and land use data, ending up with BSU roughly corresponding to plots or pixels. To sum up, results appear to be given at the level relevant for policy makers, either the study area as a whole or by administrative delineations. Moreover, BSUs of the model are seen as a purely technical aspect. Their size and nature depend on the modelling framework used, leading to large variations in size and number. Models based on a GIS often use small pixels or plots, while LUTI models focus on custom TAZ based on census blocks. Large-scale projects seem to use more disaggregated BSU, but this tendency remains unclear. It may reflect the larger capabilities of large MPO (as outlined by Lee, 2009; see section 2.2).

2.4.2 Applications of LUTI models in the EU

Data collection

This second case study is based on papers published in peer-reviewed journals with impact factor before December 2014. All other forms of communications (proceedings, working papers or chapter in books) are excluded. Three bibliographic databases have been used (Google Scolar, Scopus, and ResearchGate) and the selection rely on author's appraisal, i.e. is considered as LUTI's papers an article where the authors affirm to use such type of model. The sample obtained is limited to 19 papers, confirming the tendency highlighted by Wegener (2011a). Table A.5 summarizes the spatial characteristics of the 25 case studies presented in these papers (see also Figure 2.7). The variables collected for each publications are described in Table 2.7.

Field	Definition
Study Area	The city or region modelled
Authors	Name of the author(s)
Date of publication	Year when the paper was published
Journal	Journal in which the paper was published
Size	Mention of the size of the study area (Yes/No)
Population	Inhabitants within the study area
Area	Extension of the study area, in km^2
Map	Is a map of the study area provided (Yes/No)

Table 2.7 – LUTI model applications in Europe (bibliometrical information collected for papers published in peer-reviewed journals; results are given in Table A.5)

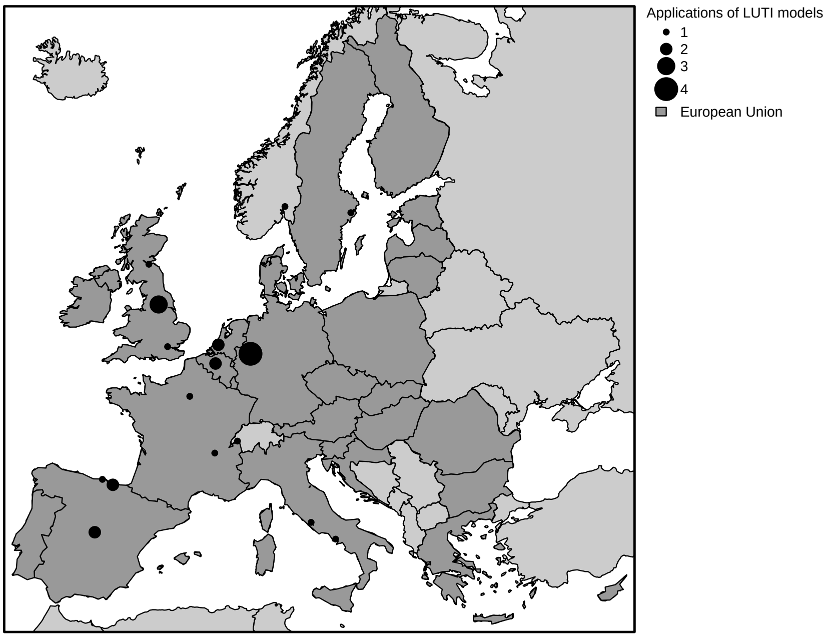

Figure 2.7 – LUTI model applications in Europe (according to papers published in peer-reviewed journals)

Results

Five case studies (Dortmund, Leeds, Brussels, Bilbao, and The Netherlands) have led to more than one publication, reducing the number of different cities or metropolitan area to 17 (Figure 2.7). This repetition can be explained by the profile of the author: academics are more inclined to publish their work than consultant. It also reflect, however, scale economies: since the data collection step is often long and difficult (Wagner and Wegener, 2007; Nguyen-Luong, 2008; Wegener, 2011a), available databases are re-used for when the original model is refined. Most case studies (23/25) are isolated urban areas (the remaining two, Zondag and de Jong, 2011 and Zondag et al., 2015, cover The Netherlands). The delineations of these urban areas are, however, not well documented. Official extent of the metropolitan area are often used, but most authors do not mention their exact limits, nor include a map devoted to the description of the study area. The choice of such study areas seems governed by policy/administrative reasons, by the agencies supplying the data, and/or by the researchers assembling the available data. No critical viewpoint appears on these choices (only Mackett, 1990 questions that problem, without solving it), neither than on their consequences in terms of urban disparities. Large variations of scale are observed, with a population within the study area comprised between 0.3 and 14 millions of inhabitants (note that 25% of the papers does not mention this simple indicator). Hence, the nature of the studied area is different from one city to another. It sometimes includes rural landscape (e.g. Paris) or catchment area of other cities (e.g. Brussels). Both factors limiting the comparability of the outputs. The questions of intercity relationships as well as the representation of “the rest of the world” while modeling urban transport/land use realities is not considered either.

2.4.3 Space in operational applications of LUTI models

The two case studies presented here are not exhaustive. Many applications of LUTI models, in the US, in Europe and, a fortiori, in the rest of the world, have not been reviewed. The fact that the same conclusion can be drawn from these two very different samples is, however, noteworthy. Spatial choices made in operational applications of LUTI model appear to be default choices, i.e. to rely on official administrative delineations or on data availability without questioning the relevance of the study area and/or of the BSU. In particular, even if LUTI models left the modellers free to select the study area on which they are applied (see also section 2.3), most applications rely on an aggregation of administrative units (counties in the US, various for Europe) that intersect the footprint of the metropolitan area. The constraint faced by the modellers, in particular the MPO’s area of responsibility, should however be recognised. Nevertheless, this lack of awareness for spatial biases constitutes, to our point of view, a clear weakness for the robustness and comparability of LUTI model's applications.

2.5 Spatial bias and LUTI models

Biases arising from the use of geographical data are not a new issue in statistical analysis. Various handbooks on quantitative geography, spatial analyses or econometrics have extensively described them, as well as the methods proposed to control or reduce their influence (e.g. Anselin, 1988b; Haining, 1993; Fotheringham et al., 2000b; Fischer and Getis, 2010; and LeSage and Pace, 2009). Spatial, in some sense, is special (Anselin, 1989). Nevertheless, the three previous sections, on the history of LUTI models (section 2.2), their internal principles (section 2.3), or their operational applications (section 2.4), suggest that space is never seen as a constraint in LUTI models. As a result, the effects of spatial bias on LUTI models' behaviour and outputs remain largely unexplored, and probably underestimated. This section aims at formalising their potential influence on LUTI models, first by presenting the three main theoretical and conceptual framework of spatial bias, then by reviewing their relevance for the spatial extent and resolution.

2.5.1 Theoretical and conceptual background

The modifiable areal unit problem

The well-known Modifiable Areal Unit Problem (hereafter MAUP) attempts to understand why identical individual data yield to different statistical results when aggregated in varying ways. The first evidence of this issue is found in Gehlke and Biehl (1934), although its definition was introduced by Openshaw and Taylor (1979) and Openshaw (1984). The MAUP is classically divided between two components: (1) the *scale effect* deals with the variations of statistical measures with the size of the areal unit (Fotheringham and Rogerson, 2009; Briant et al., 2010; Arbia and Petrarca, 2011). Note that for a finite number of observations, aggregation of individual data into areal units of increasing size leads to a decrease of the number of observations. The (2) *aggregation effect* is defined as the variations of statistical measures arising from the aggregation into areal units based on a different criterion (Arbia and Petrarca, 2011). In this case, the number of the areal units remains identical, but their borders are different.

The *aggregation effect* is often found to lead to random variations, and is largely considered as intractable (Fotheringham and Wong, 1991). Regarding the *scale effect*, Openshaw (1984) suggested four initial solutions to control and/or reduce its influence. First, one could simply ignore the MAUP and hope that the outcome of the research is still significant. Second and third solutions are similar: they aim at reducing the effects of the MAUP by using either individual data, or the "appropriate scale of analysis". The last solution proposed by Openshaw (1984) is to draw the spatial units so that the analysis produces a predicted outcome. None of these solutions can, however, be easily applied to LUTI models. Appropriate scale of analysis or custom spatial units are indeed difficult to use in the field of LUTI models, since data are most likely to be available only for administrative units, and these units remain the relevant level for policy makers.

First studies on the MAUP focused on correlations (e.g. Gehlke and Biehl, 1934; Openshaw and Taylor, 1979; and Openshaw, 1984), but were soon extended to multivariate regression analysis. The literature on that particular topic is abundant (e.g. Arbia, 1989; Fotheringham and Wong, 1991; Amrhein, 1995; Reynolds and Amrhein, 1998; and Arbia and Petrarca, 2011). These work show a systematic variation of parameter estimates when the number of zone decreases, some becoming more negative and other becoming more positive. The influence of the MAUP on parameter estimates of multivariate regression methods is often considered unpredictable (Arbia, 1989; Fotheringham and Wong, 1991). More recent works, however, have suggested a link with the spatial autocorrelation structure of the variables (Reynolds and Amrhein, 1998). A positive autocorrelation means that adjacent areal units have similar values. Parameter estimates are less sensitive to the MAUP in that situation. Arbia and Petrarca (2011) reaches similar conclusions for a spatial auto regressive model. Note that these results apply only to the classical definition of the MAUP. To our knowledge, the influence of the size of the study area on linear regression's parameter estimates remains unexplored.

The sensitivity of Discrete Choice Models (from now on DCM) to the MAUP has been far less studied. Arauzo-Carod and Antolín-Manjón, 2004 have compared firm's location choices for three levels of administrative units, using both DCM and Count Data Models (CDM). They observe significant differences between parameter estimates and conclude that location choice factors do not act uniformly with the scale over broad geographic regions. However, their study area (Catalonia region, Spain) and areal units (provinces or municipalities) cannot be compared easily to typical applications of LUTI models (metropolitan areas with small BSU such as census tract; see chapter 2). The availability of detailed data sets on firm's locations means that empirical studies of households or employment location choices have evolved from aggregated to disaggregated areas (e.g. census wards instead of municipalities or regions, see e.g. McCann and Sheppard, 2003; Guimaraes et al., 2004; and Arauzo-Carod et al., 2010). An example for residential location choice in Paris (France) can be found in de Palma et al. (2007): the conclusion is that the factors driving these choices vary with the size of the spatial unit considered (municipalities or grid cells of 500 by 500 meters), but the paper does not provide a complete analysis on the influence of the MAUP.

Regression analysis and discrete choice models are the most used econometric methods in LUTI models (see chapter 2). Hence, for further details on the influence of the MAUP on statistical analysis, we refer the reader to above-mentioned handbooks. Note that the sensitivity of a spatial interaction gravity model (a procedure commonly found in older aggregated LUTI models) to the MAUP can be found in Arbia and Petrarca (2013).

Note that the ecological fallacy problem (Robinson, 1950) is closely related to the MAUP, even if the two concepts are presented separately in the literature. It states that correlations between individuals can be different from the one computed at the level of the group to which the individuals belong. This issue arises when the characteristics of the individuals are deduced from aggregated data. Hence, ecological fallacies do not affect estimations of econometric methods *per se*, but rather data preparation and interpretation. A typical example for LUTI models is that synthetic population often has to be generated based on aggregated marginal controls (e.g. Ye et al., 2009; Farooq et al., 2015a; Farooq et al., 2015b).

The uncertain geographic context of problem

Spatial bias can also be linked to the less-known Uncertain Geographic Context of Problem (UGCoP, see Kwan, 2012). It states that the delineation of areal units used as observations may not correspond to the true causally relevant units (Kwan, 2012). It differs from the MAUP in the sense that it does not assess the influence of various delineations of areal units, but aims to identify the true delineations relevant for the study of a given phenomenon. The examples given by Kwan (2012) are focused on the field of health geography, and no systematic sensitivity analysis appear to exist. Nevertheless, in the field of LUTI models, the UGCoP can be related to the question of the neighbourhood taken into account by households to select their residential location. Guo and Bhat (2004) show that it can be greater than administrative delineations (census block). Hence, households may consider the land use and amenities around their potential dwelling within a radius larger than the areal units used in the model. Note that the shape of the neighbourhood taken into account will also affect parameter estimates of a residential location choice model (see Guo and Bhat, 2007), raising the question of the definition of the true neighbourhood considered by agents .

The definition of the study area is also closely related to the UGCoP issue. LUTI models are generally applied on metropolitan areas (see chapter 2). It is a known issue that due to the sub urbanisation process, urban areas do not have a clear border anymore (see Cheshire and Gornostaeve, 2002; Dujardin et al., 2007; Cörvers et al., 2009; or Farmer and Fotheringham, 2011). Finally, not the that no methodology has been proposed yet to solve the UGCoP, although it most likely involves a better comprehension of individual's behaviours (Kwan, 2012).

2.5.2 Relevance of spatial choices for LUTI models

The general lack of awareness for spatial biases observed in the field of LUTI models means that the spatial issues identified since the beginning of LUTI models remain largely unsolved (Lee, 1973, te Brömmelstroët et al., 2014). Let us briefly go back to the seven sins of Lee (1973). The *grossness* issue has seen improvements, thanks to the higher level of detail of current models that allows simulating more policies. Nevertheless, the case studies reviewed in section 2.4 show that results of LUTI models are still presented at an aggregated level. The *hungriness* is perhaps even more prevalent for micro simulation models (see Wegener, 2011a) than for models derived from Lowry (1964). Finally, the *wrongheadedness* is still a fundamental issue, especially since different assessments based on LUTI models' outputs (e.g. environmental impact) have become, in some cases, legally binding (see section 2.2 and 2.4.2). Overall, both *spatial extent* and *spatial resolution* have implications for LUTI models developments and applications.

Spatial extent

The footprint of a city is a complex phenomenon, driven by multiple factors such as its geography, its history, the local governance process and actors, and its size and interconnections, relative to other cities (Rémy, 2004; Paulet, 2009). Many operational applications of LUTI models are publicly funded (see section 2.4), which may retrain the modellers to use the jurisdiction of these public authorities as study area. Wether this is a real constraint or a default choice remains, however, an open question. In any case, the *spatial extent* issue is closely linked to the delineation of cities' catchment area, that has fascinated geographers for decades (see chapter 1 and chapter 6). It has been largely demonstrated that cities sprawls out of their administrative boundaries, but the relevant delineation of a city depends on the goal of the study (Dujardin et al., 2007). The lack of interest observed in the definition of the study area raises, therefore, the risk of biasing the estimation of the LUTI model.

In this chapter, we show that the *spatial extent* chosen may affect both the data collection and processing step, as well as the behaviour of the model. Two issues should be particularly highlighted. First, if the study area covers multiple administrative entities, there is a risk that some data will not be available and/or will have different definitions. For instance, land use regulations are often a regional competence meaning that land use categories can vary across the study area. This potential bias is, however, specific to each case study. Making more formalisation or generalisation is, therefore, impossible and it will thus not be assessed in this thesis. The second issue, i.e. the influence on the behaviour of the model, can be related to the Uncertain Geographic Context of Problem. In particular, the *spatial extent* chosen for the model may include rural areas and/or part of other cities. This will increase the heterogeneity of the study area, potentially affecting the goodness-of-fit of the model for both econometric sub models (which will be explored in chapters 3 and 4) and the model outputs (see chapter 5).

Spatial resolution

The areal units (or BSU) used in a LUTI model are often seen as a purely technical aspect (see section 2.4). Modellers select one model and apply it using the areal units for which data are available. There is often no integration between the land use and transport components of the LUTI model, resulting in a dichotomy between land use and traffic analysis zones. In early applications of LUTI models, it was also frequent for an aggregated traffic model to be combined with land use model based on raster data. The choice of the areal units is, nevertheless, likely to influence both the data collection and processing steps and the behaviour of the model system. Leading, potentially, to variations in its outputs.

The first aspect, data collection and preparation, relates to spatial aggregation issues (see Goodchild and Gopal, 1989). The variety of the spatial data required by LUTI models means that they will often be maintained by different providers. For instance official census data for socio-economic characteristics but regional survey for travel behaviour. Land-use regulations are also likely to involve different political bodies. In Belgium, land use planning is the responsibility of the regions, but municipalities can define additional rules. The use of data available for different spatial units raises the risk of ecological fallacy (Robinson, 1950). As for the *spatial extent*, however, the presence of this bias depends on the case study. Its influence being impossible to formalise, it will not be assessed in further analyses.

Even when the database is completed, the size of the areal units may still affect the behaviour of LUTI models. The *spatial resolution* issue is indeed conceptually identical to the *scale effect* of the well-known Modifiable Areal Unit Problem (Openshaw and Taylor, 1979). Section 2.3 shows that LUTI models heavily rely on econometric method, especially regression analysis and discrete choice models. Parameters estimates of both methods are known to be sensitive to the MAUP. The forecasts of both future real estate prices and future location of agents can, therefore, affected by the *spatial resolution* of the model system. The direction and magnitude of this potential bias in LUTI models has, to our knowledge, never been assessed. Hence, chapter 3 will explore the sensitivity of discrete choice models to this *spatial resolution* issue, and chapter 4 and chapter 5 those of the model's outputs.

2.5.3 Conclusion

To our opinion, limiting factors (administrative regulations or data availability constraints) doesn't excuse the lack of interest for the *spatial extent* and *resolution* of the model. The current debate on the optimal level of disaggregation of micro-simulation models (see Wegener, 2011a) is certainly a step in the good direction. Its emergence, however, is partially linked to the failure of large-scale disaggregated LUTI models' project (e.g. Wagner and Wegener, 2007; Nguyen-Luong, 2008). Knowing that spatial biases may be present should encourage the realisation of a careful exploratory spatial data analysis, to assess their potential influence, instead of ignoring it. The main challenge of today's LUTI model, as we intend to demonstrate in this thesis, is thus not to increase the level of details. As pointed out by Wegener (2011a), it would rather be to identify the optimum level of conceptual, spatial, and temporal resolution for each component of the model. There is a need for multi-level models where the level of details can be adjusted to the process simulated, and to data availability (see chapter 6). There is also, and perhaps even more, a need for a better awareness of spatial biases in LUTI models applications

Part II

Sensitivity of LUTI models' econometric components

CHAPTER 3

Boundary effect on land price determinants

3.1 Introduction

The aim of the chapter is to extent the state-of-the-art on the influence of spatial biases on regression analysis, by examining their sensitivity to the *boundary effect*, i.e. variations of the size of the study area. Contrary to the *scale effect*, this issue of the size and composition of the study area has received no or little attention (see chapter 2). Although the use of regression procedure in state-of-the-art LUTI models is limited, chapter 2 shows that *UrbanSim* (among others) relies on such methods to forecast the evolution of real estate prices. For consistency with the general topic of the thesis, the research question of this chapter is thus the following. How does the functional delineation of the city influences land prices determinant?

A sensitivity analysis of (developable) land price to the delineation of the study area is conducted on the case study of Brussels (Belgium). A hedonic model (log-linear and semi-parametric specifications) is estimated for 12 delineations of the city (morphological, functional, administrative, or based on transportation infrastructure), as well as on automatically constructed delineations. The underlying assumption, grounded in the classical bid rent model of Alonso (1964), is that the land price is a function of the accessibility to the CBD. The Alonso-Muth model assumes a clear-cut limit of cities, located where the opportunity costs of urban and agricultural land-use are equal. Hence, in rural areas the influence of accessibility to the CBD on land prices is expected to be null. Assuming that local amenities have been controlled for, the effect of the accessibility to the CBD on land prices should thus decrease when the size of the delineation increases.

The emergence of sub centres, however, leads to non-linearity in the decrease of land price when the distance to the CBD increases (see e.g.McMillen, 1996; Anas et al., 1998; Ahlfeldt, 2011). Fujita and Ogawa (1982) and Fujita (1989), among others, have proposed analytical model to introduce such sub centres in the general Alonso-Muth model. Another limitation of the model is that suburban settlements do not show a dichotomy between urban and rural land uses. Nevertheless, Cavailhès et al. (2004) have proposed an analytical model where households commuting to the CBD and farmers can be present in a mixed belt surrounding city centre. Hence, this chapter aims to test the intuition that the influence of accessibility to the CBD on land prices will be affected by local conditions (secondary centres or rural areas having loose relationship with the CBD) and that these variations allow to assess the efficiency by which the influence area of a city is captured by different delineations of that city. This work can be related to Bode (2008), who proposed to use the fraction of land prices attributable to economies of urban agglomeration to delineate cities. The bottom-line of this chapter is that different delineations can be proposed for the same city, but that they have different meaning and are not interchangeable, nor easily comparable.

This chapter is organised as follows: section 3.2 presents the case study and section 3.3 the methodology. Results are presented in section 3.4, and discussed in section 3.5. Finally, Section 3.6 summarises the implications for operational applications of LUTI models.

3.2 The case study

Brussels is the capital and main employment center of Belgium. It is an outlier among European cities in that its official extension (the Brussels-Capital Region, hereafter BCR) is so small, yet so politically significant, relative to its functioning whole (Thisse and Thomas, 2007; Cheshire, 2010). This BCR accounts for 1.1 million inhabitants, while the total metropolitan area can reach, depending on its definition, up to 2.5 millions inhabitants (see Table 3.1). In terms of employment, 18.8% of Belgian GDP was produced in the BCR in 2007, and about 670 000 jobs were located in it, 229 500 being occupied by people commuting daily from Flanders, plus 126 500 from the Walloon Region (Thisse and Thomas, 2010). In the BCR, there is *"a socio-economic cleavage, with working class neighbourhoods along a north to south - west axis, and a rich south - east quadrant"* (Vandermotten et al., 2010; pp. 82). This deprived axis

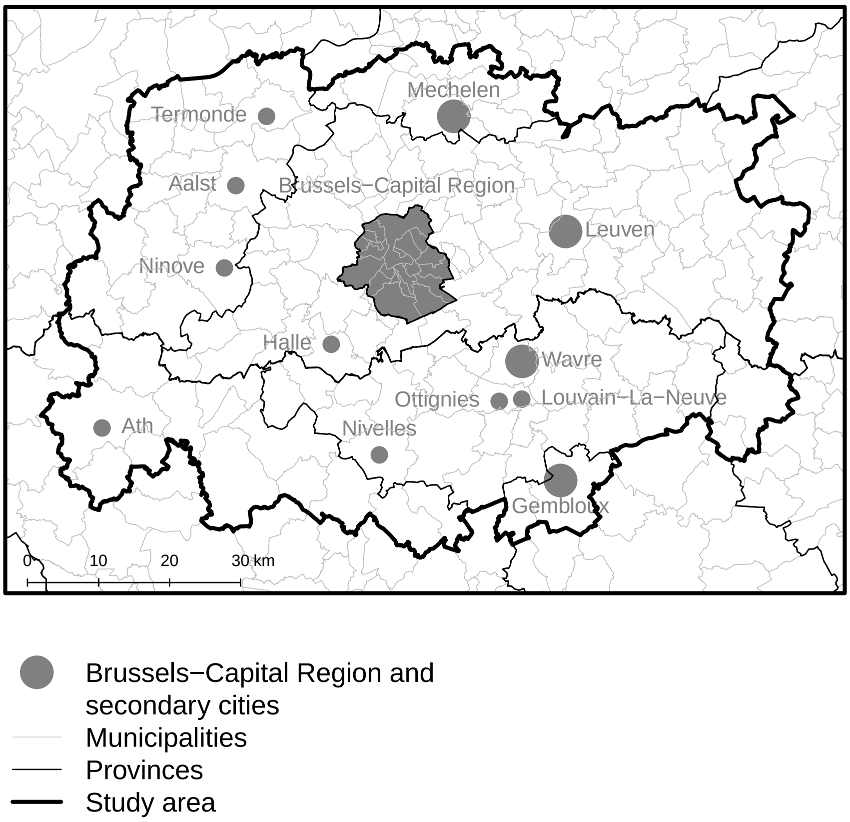

Figure 3.1 – Study area

goes beyond the boundaries of the BCR (see e.g. Thomas and Zenou, 1999; Kesteloot et al., 2001; Dujardin et al., 2008) and is visible in the distribution of land prices and median annual income (Figure 3.4).

Different secondary cities are found in the vicinity of Brussels (Figure 3.1). The main ones are Leuven and Mechelen (both in Flanders) and the conurbation of Wavre, Ottignies, and Louvain-La-Neuve in Wallonia. Compared to the Belgian' average, real estate prices are high in the entire study area, but mostly in Flanders (see section 3.2.2).

3.2.1 Definitions of Brussels' catchment area

Given the political context of Belgium, none of the functional delineations proposed for the Brussels metropolitan area have made their way out of the academic world (publicly funded researches on Brussels are often limited to the BCR). This lack of consensus has encouraged geographers and economists to propose their own definition of the Brussels metropolitan area, making this city a fascinating case study for comparing urban delineations. The different delineations of Brussels used in this work are presented hereafter.

Delineations proposed in the literature

Two administrative delineations are used (Figure 3.2a). As already mentioned, the official administrative extent of Brussels is the BCR. Before 1995, it was included in the Province of Brabant, which can thus also be considered as a potential delineation of Brussels:

- The **Brussels-Capital Region** (BCR) includes 19 municipalities and corresponds to the very dense urban center;
- The **Former province of Brabant** (Brabant)[1] is constituted of 111 municipalities

Functional delineations of Belgian cities were first proposed in 1979. Latest revision was made by Van Hecke et al. (2009), using the population census of 2001. It combines a morphological agglomeration, defined by the contiguity of built-up areas (maximal distance between buildings less than 200 meters), with three nested delineations that match the boundaries of the municipalities (the smaller administrative units for which land prices are available, see section 3.3). These latter three have thus been considered (see below). Figure 3.2 shows the extension of these delineations for Brussels. Note that a map of all urban regions in Belgium is given in appendix (Figure B.2).

- The **Operational Agglomeration** (OA) is the set of municipalities having more than 50% of their inhabitants in the morphological agglomeration. In 2001, it includes 36 municipalities, and corresponds to the densely build area;
- The **Urban Region** (UR) (62 municipalities) includes the OA plus the suburbs (26 municipalities), which are defined by a mix of socio-economical indicators. UR is widely used as definition of the Brussels city region in scientific work (e.g. Riguelle et al., 2007; Verhetsel et al., 2010);
- The **Metropolitan Labour Area** (MLA) includes 122 municipalities: the UR plus the 60 municipalities of the Commuter Living Zone, or CLZ,

[1] Former since it has been divided among the BCR, the Walloon Brabant and the Flemish Brabant provinces in 1995

defined as all municipalities having more than 15% of active population commuting daily to the OA. Note that the presence regional cities (Mechelen and Leuven) on north and east of Brussels limit the extension of the MLA on these directions (Van Hecke et al., 2009).

Sensitivity analysis of these functional delineations to the threshold values used can be found in Dembour (2004) and Dujardin et al. (2007). However, these works have not been considered here, since the differences with the delineation proposed by Van Hecke et al., 2009 are limited. Transport infrastructures, such as the RER network, have also been used to delineate Brussels. The RER, for "Réseau Express Regional", is a fast train network to and from Brussels, currently under construction. The extension of this RER network has been defined by law (Figure 3.2c):

- The **Official RER Zone** (RER) is defined as all municipalities *"within a radius of about 30 km from Brussels"* (Moniteur Belge, 2004; pp. 97), and includes 136 municipalities;
- An **Inner RER Zone** (InnRER) is defined by the same law as the OA plus all municipalities contiguous to the BCR, for a total of 41 municipalities;
- Furthermore, potential effects of the RER have been studied using an **Extended RER Zone** (ExtRER) of 147 municipalities (the RER Zone plus 21 additional municipalities, without justification; see Boon and Gayda, 2000).

Different works have aimed to apply innovative methodologies (fractal geometries and clustering of OD-matrices) to delineate Brussels (Figure 3.2d and e). These methods are original in the sense that the delineation of the city is endogenously determined, rather than based on exogenously fixed threshold values. The resulting delineations of Brussels are:

- A **fractal morphological delineation** (Fractal), using the dilation of individual building footprint method proposed by Tannier et al., 2011. Roughly, the urban boundary corresponds to the point of maximum curvature on the dilation curve, i.e. the curve of the number of built clusters as a function of the buffer width used in the dilation step. This morphological delineation intersects 48 municipalities (see Tannier and Thomas, 2013);
- A **phone basin** (Phone, see Blondel et al., 2010) based on an Origin-Destination (OD) matrix of mobile phone communications (at the municipality level and for the entire country). A network partition methodology based on the modularity of the graph (Blondel et al., 2008) is used

Delineations	n	Population million	Area km^2
Brussels-Capital Region (BCR)	19	1.039	162
Operational Agglomeration (OA)	36	1.441	577
Inner RER zone (InnRER)	41	1.533	726
Fractal morphological agglo. (Fractal)	48	1.576	932
Urban region (UR)	62	1.820	1 527
Phone basin (Phone)	66	1.731	1 718
Job basin (Job)	105	2.197	3 282
Former province of Brabant (Brabant)	111	2.467	3 376
Metropolitan labour area (MLA)	122	2.662	4 153
Official RER zone (RER)	126	2.953	4 151
Extended RER zone (ExtRER)	147	3.267	4 969
Union (Union)	160	3.382	5 602

Table 3.1 – **Delineations of Brussels** (values for 2010; n = number of municipalities included in the delineation)

to draw communities of municipalities. The phone basin around Brussels encompasses 66 municipalities;

- A **job basin** (Job, see Thomas et al., 2013) that uses the same methodology than phone basins but on an OD matrix of home-to-work commuting fluxes in 2008. The methodology applied lead to a job basin of 105 municipalities, centred on Brussels.

Finally, the Union of the delineations of Brussels is here defined as all municipalities (160) belonging to one of the aforementioned functional delineation. Table 3.1 and Figure 3.2 summarise the extension of these delineations of Brussels.

Automatically generated delineations

In order to study the variations of land price determinants independently from the underlying urban structure, a continuum of delineations has been generated, by the following procedure:

1. The BCR (19 municipalities) is the initial delineation of the continuum (iteration 0), and the Union of delineations (160 municipalities) as its maximal extension;
2. Delineation n is defined as all municipalities belonging to the $n-1$ delineation, plus the municipality not included in $k-1$ having the minimal Euclidian distance to the centroid of the municipality of Brussels;

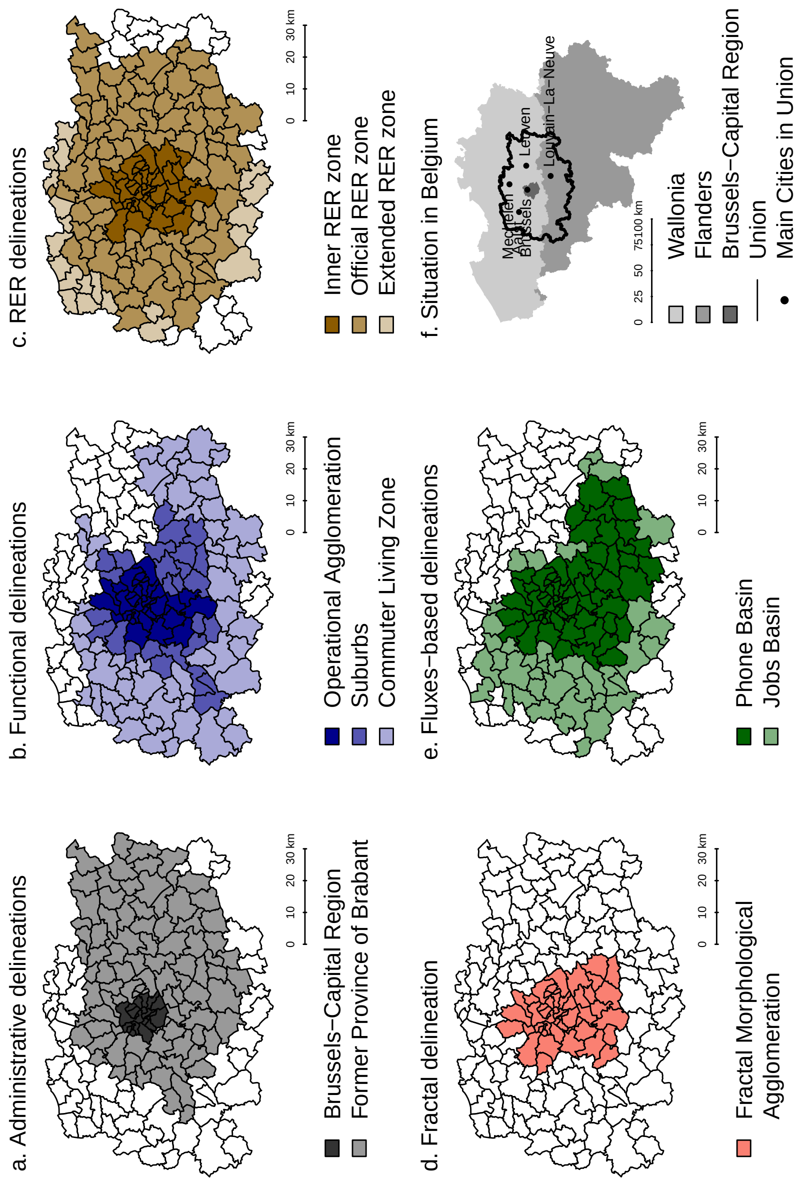

Figure 3.2 – Delineations of Brussels.

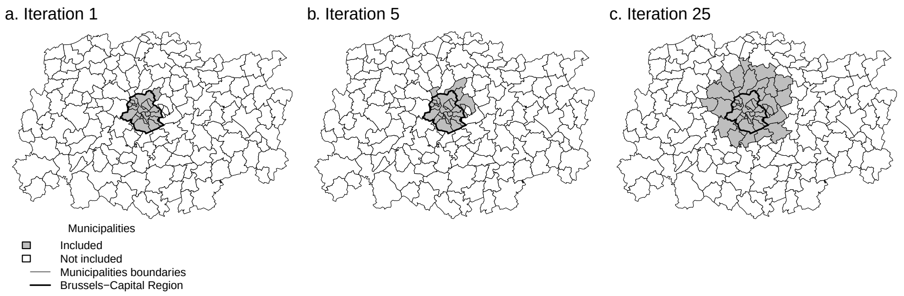

Figure 3.3 – Continuum of delineations of Brussels (note: complete extent = Union)

3. This procedure is repeated iteratively (Figure 3.3) until that all municipalities are included. The continuum thus consists in 141 delineations (i.e. 160 - 19).

It should be noted that this continuum assumes that the influence of Brussels spreads evenly in all directions, which is not the case due to the presence of regional cities close to Brussels (see section 3.5). It is, therefore, less meaningful than functional delineations.

3.2.2 Data

The aim is here to compare different delineations of Brussels, not to build the best possible model of land prices for each of them. For estimations of detailed land price models on either Belgium or Brussels, the reader can refer to Goffette-Nagot et al. (2011), Cavailhès and Thomas (2013), or Pholo Bala et al. (2014). We restrain ourselves here to a simple specification, with a limited amount of explanatory variables. This approach has been preferred due to the specificities of LUTI models (see below), but also because informations on land prices are rather limited in Belgium (no individual transactions are available). The only data disclosed by the Belgian Directorate-general Statistics and Economical Information (or DGSIE) are an annual series, at the level of the municipalities, of the number of transactions per annum, the overall value of these transactions (€) and the total surface of the parcels sold (in square meters). Hence, the average price for 1 square meter of developable land sold within the municipality between 2006 and 2008 is used here (Table 3.2). A three-years period reduces biases due to very small numbers of transactions (a map of the total number of transactions per municipality is given in appendix, see Figure B.1). Still, five municipalities within our study area have less than 10 transactions and have been excluded from the analysis. Prices are deflated between years by the consumption price index. Developable land is used as a proxy for real estate prices. Since they have fewer intrinsic characteristics, it can be assumed that their average price depends mostly of the municipality location. As expected, land prices are maximal for municipalities close to Brussels (Figure 3.4) and decrease with the distance. All other things being equal, prices also tend to be higher in Flanders than in the Walloon Region, due to population densities and to land use policy (Goffette-Nagot et al., 2011).

The control variables reflect these limitations of the dependant variable. They can be divided into three groups: socio-economic, local amenities, and accessibility indicators. Variables of the first group are based upon the Alonso-Muth model. For the second group, we rely on a subset of the variables tested by Goffette-Nagot et al. (2011), while the third group includes tow accessibility indicators of different nature, as in Ahlfeldt (2011). Note that the use of simple variables is consistent with the specificities of LUTI models. Residential units' attributes are typically not represented, forcing the modeller to rely on zonal characteristics to forecast real estate prices (see chapters 6 and 7). Moreover, the number of zonal characteristics whose evolution is forecasted by the model is limited, which also constrain the modeller towards simple specification (see Nguyen-Luong, 2008). The limited number of independent factors considered here constitutes, therefore, a weakness of this work, but a weakness that will be present as well in most operational applications of LUTI models.

Hence, these control variables are the following: Population density, which is the number of inhabitants per square kilometres in each municipality (Figure 3.4) in 2008, and Income, equal to the median earnings per household in euros for fiscal year 2008, are used as socio-economic indicators. Sources for these data are the DGSIE. The Travel time to the closest main (IC/IR) railway station[2] is used as a local amenities' indicator and accounts for the proximity of the municipalities to both the public transport network and to secondary centres with retail activities. Forest cover, expressed as the share of the total surface of the municipality covered by forest (from CORINE land-cover database, see EEA, 2006), is used as green amenities' indicator. Given the average size of municipalities (34 km^2), it is impossible to test for the presence of parks or schools (Goffette-Nagot et al., 2011). Finally, a dummy variable (Wallonia) taking the value 1 if the municipality belongs to the Walloon region, and 0 otherwise (BCR or Flanders) is considered.

For comparing the different delineations of Brussels, the key variable of the model is an indicator of the accessibility to Brussels, since it will capture the influence of the city on its neighbouring areas. For that purpose, the Travel time to the centroid of the municipality of Brussels by car, in minutes and congestion included, is used (data from Vandenbulcke et al., 2007). In addition,

[2]IC (InterCity) and IR (InterRegio) is the name given by Belgian railways for fast, direct trains. By extension, the « IC/IR stations » are the stations where these trains stop.

Variable	**Units**	**Mean**	**SD**	**Min**	**Max**
Land prices (transactions)		110	70	9	452
Land prices (value)	€/m^2	169.1	181.6	28.1	1 664.9
Population density	hab/km^2	1 601	3 474	106	20 630
Median annual income	× 1000 €	20.9	2.2	12.3	32.1
Forest cover	%	6	9	0	56
Time to IC/IR stations	minutes	5.6	3.7	0.1	17.2
Travel time to BXL	minutes	43.1	17.6	1.0	79.6
Accessibility to jobs		10.52	0.24	10.10	11.27

Table 3.2 – **Variables of the land prices model** (for the Union delineation; SD = standard deviation)

a gravitational measure of the Accessibility to jobs (A_i) in each municipality i is also used.

$$A_i = log(\sum_{j=1}^{n} d_{ij} \times w_j) \tag{3.1}$$

In equation 3.1, d_{ij} denotes the travel time by car between the centroid of the municipality i and the centroid of the municipality j (data from Vandenbulcke et al., 2007) and w_j the number of jobs located in j. Note that all municipalities in Belgium are taken into account to compute A_i. Job' data come from the Belgian National Social Security Office (ONSS, 2015), and exclude self-employees. Although it has been shown that the linguistic border between French and Dutch-speaking part of Belgium significantly reduces interactions (Dujardin, 2001; Blondel et al., 2010), estimating the magnitude of this effect is a complex task. Hence, we do not consider any border effect here.

Given the importance of the BCR as an employment centre, the correlation (Pearson product-moment) between the travel time to the Brussels CBD and the Accessibility to jobs is high (-0.86***). Still, their nature is different: the Travel time to the Brussels CBD only measures the accessibility to that specific location, while the Accessibility to jobs takes into account all potential destinations. Hence, it is here expected that the effect of this latter variable should be similar for all delineations of Brussels, while the influence of the travel time to the Brussels CBD would vary according to the size or the composition of the delineation used.

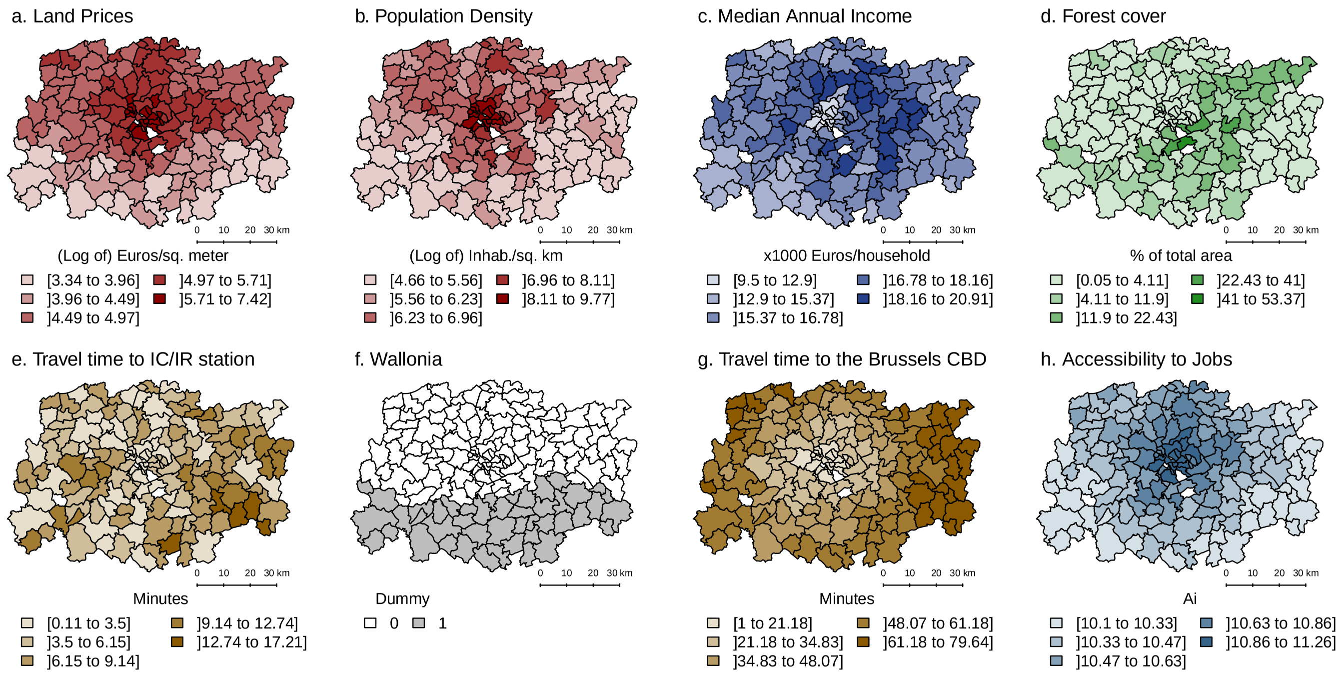

Figure 3.4 – Variables of the land prices model (note: blank municipalities = no data for land prices)

3.3 Methodology

The influence of the delineation of Brussels on land price determinants is assessed using a two steps methodology. First, a hedonic model of land price is estimated on the functional and automatically - generated delineations of the city (see section 3.2). This model is estimated by OLS (equation 3.2) and by a semi-parametric specification, as in equation 3.3 (mixed model representation of penalised splines, see Ruppert et al., 2003). For the latter one, the relationship between land prices and the accessibility indicator (either the travel time to the Brussels CBD or the accessibility to jobs) is not constrained to a linear form.

$$Ln(P_i) = \alpha + \beta_j X_{ij}^{control} + \beta_a X_{ia}^{accessibility} + \epsilon_i \tag{3.2}$$

$$Ln(P_i) = \alpha + \beta_j X_{ij}^{control} + f(X_{ia}^{accessibility}) + \epsilon_i \tag{3.3}$$

Where P_i is the average selling price, in euro per square meter, of developable land in the municipality i, $X_{ij}^{control}$ a vector of the j control variables in i and $X_{ia}^{accessibility}$ the accessibility indicator a in i, α is a constant and ϵ_i the random, normally distributed, error term associated to each observation i. Definition and sources of the dependent and independent variables have been discussed in section 3.2. Since this work is exploratory, the benchmark specification of the land price model for each delineation of Brussels has been determined by a stepwise procedure, using the Akaike Information Criterion to select the set of control variables (see Venables and Ripley, 2002). Three different cases are considered, in order to compare the effects of the accessibility indicators (Table 3.3). The specification implemented for the continuum of delineations is based on the variables found significant for functional delineations (see section 3.4).

Note that we did not measure the presence of spatial auto correlation between the dependant variable nor than between the residuals of the regression. Its presence is very likely given the size of the BSU (see Goffette-Nagot et al., 2011). Nowadays, most estimations of hedonic prices rely thus on more advanced econometric specifications than OLS, especially spatial regression models. This option has not been considered here, since such methods are not yet implemented in *UrbanSim* neither than in other state-of-the-art LUTI models. See chapter 7 for details.

This first step allows for a global analysis of the influence of the boundaries of the study area on the land price model. In the second step, local variations of the relation between land prices and accessibility have been examined through a Geographically Weighted Regression (GWR) model (Fotheringham et al.,

Case	Control variables	Accessibility indicators	
		Travel time	Accessibility to jobs
A	All	Yes	Yes
B	All	Yes	
C	All		Yes

Table 3.3 – Stepwise regression (variables included in each specification)

2002). Its specification is given by equation 3.4, where u indicates that the parameter $\beta_k(u)$ describes the relation between land prices and the independent variable k around the location u, and is specific to that location (Charlton and Fotheringham, 2009). Other notations are identical to those of equation 3.2.

$$Ln(P_i)(u) = \alpha + \beta_j(u)X_{ij}^{control} + \beta_{ia}(u)X_{ia}^{accessibility} + \epsilon_i \tag{3.4}$$

For each accessibility indicator, three specifications are estimated. (a) A simple one with the accessibility indicator as the only independent factor, (b) a control specification, including the variables found to have an influence on land price in all cases (see section 3.4), and (c) a fitted specification using the variables selected by case B and C of the stepwise regression for the Union delineation. A Gaussian kernel is used, and the bandwidth is determined by the cross-validation optimization methods (Fotheringham et al., 2002). For the travel time to Brussels, the bandwidth is equal to 3.9 kilometres for specification (a), 7.3 for specification (b), and 23.7 kilometres for specification (c). For the accessibility to jobs, it is of 5.4, 7.1, and 114.7. The GWR procedure estimates the local value of the parameter estimates ($\beta_{ia}(u)$) for each observation i. The weights affected to all other observations j are based on their Euclidean distance with i. If j is inside the bandwidth, its weight decrease with the distance following a gaussian curve. If j is located further away than the limit of the bandwidth, it receives a null weight. The set of observations used to compute the local value of the parameter estimates is thus different for each observations, leading to the results showed by Figure 3.5.

3.4 Results

The parameter estimates of the land price models are presented in Table B.2 (in appendix). Using the variables selected by the stepwise procedure (Table B.1, in appendix), functional delineations of Brussels can be divided into two groups that correspond roughly to "small" versus "large" delineations of Brussels. Results should then be considered as illustrative for these delineations.

Moreover, it appears that the AIC criterion leads to keep into the model variables not significant at the 5% level. This is especially the case for these small delineations, and for the Forest cover variable.

As expected, Population density and Income influence land prices in all delineations. The local amenities' variables (Forest cover, Travel time to the closest railway station and the Wallonia dummy) are only selected in large delineations. When the two accessibility indicators are included in the models (case A) the Accessibility to jobs is always preferred to the Travel time to Brussels CBD. For case B, the Travel time to Brussels is only selected for large delineations. For case C, the Accessibility to jobs is selected for all delineations except the Urban Region.

The specifications estimated for cases' B and C are re-used for the semi-parametric regression model. For automatically generated delineations, the specification is identical for all iterations, and includes the variables found to be selected in all cases: Population density, Income (plus the Wallonia dummy when the continuum reaches this region) and, according to the case considered, the Travel time to Brussels or the Accessibility to jobs. For a given delineation, the parameter estimates of the control variables are never significantly different for the OLS or semi-parametric methods.

Nevertheless, Figure 3.5 shows that if the land price gradient is almost linear for the Accessibility to jobs (but become slightly steeper for low values of accessibility), this is not the case for the Travel time to Brussels. For this variable, the land price gradient is flat for the municipalities close to Brussels, become steep for those located between about 20 to 35 minutes from the CBD and then flatten again. This pattern is visible for all functional delineations, although relatively large variations are perceptible for the continuum (gray background). Hence, Figure 3.5 demonstrates the difference of nature between the Accessibility to jobs and the Travel time to Brussels, and suggests that the influence of the latter may be used in future work to assess the extension of the metropolitan area of Brussels.

A geographically weighted regression model has then been used to investigate local variations of the relationship between land prices and the Travel time to Brussels or the Accessibility to jobs. Figure 3.6 shows the variations of the parameter estimates of the accessibility indicator for the different specifications of the land price model (control specification is identical to those of the continuum). Simple and control specifications lead to counter factual parameter estimates for some municipalities (those close to Gent and Antwerp in the north, or Charleroi in the south), suggesting that they do not belong to the influence area of Brussels. This effect disappears for the fitted specification. The influence of the travel time to Brussels on land prices remains, nevertheless, minimal for these municipalities.

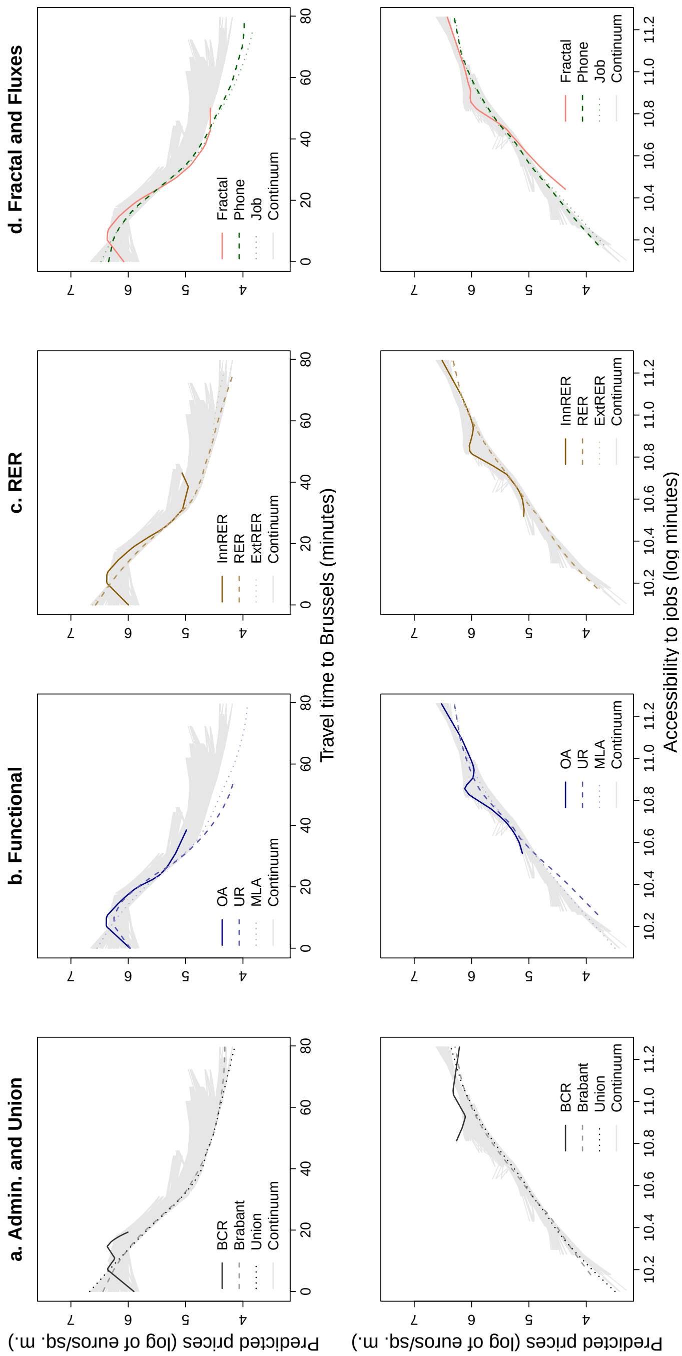

Figure 3.5 – **Predicted land prices** (semi-parametric specification)

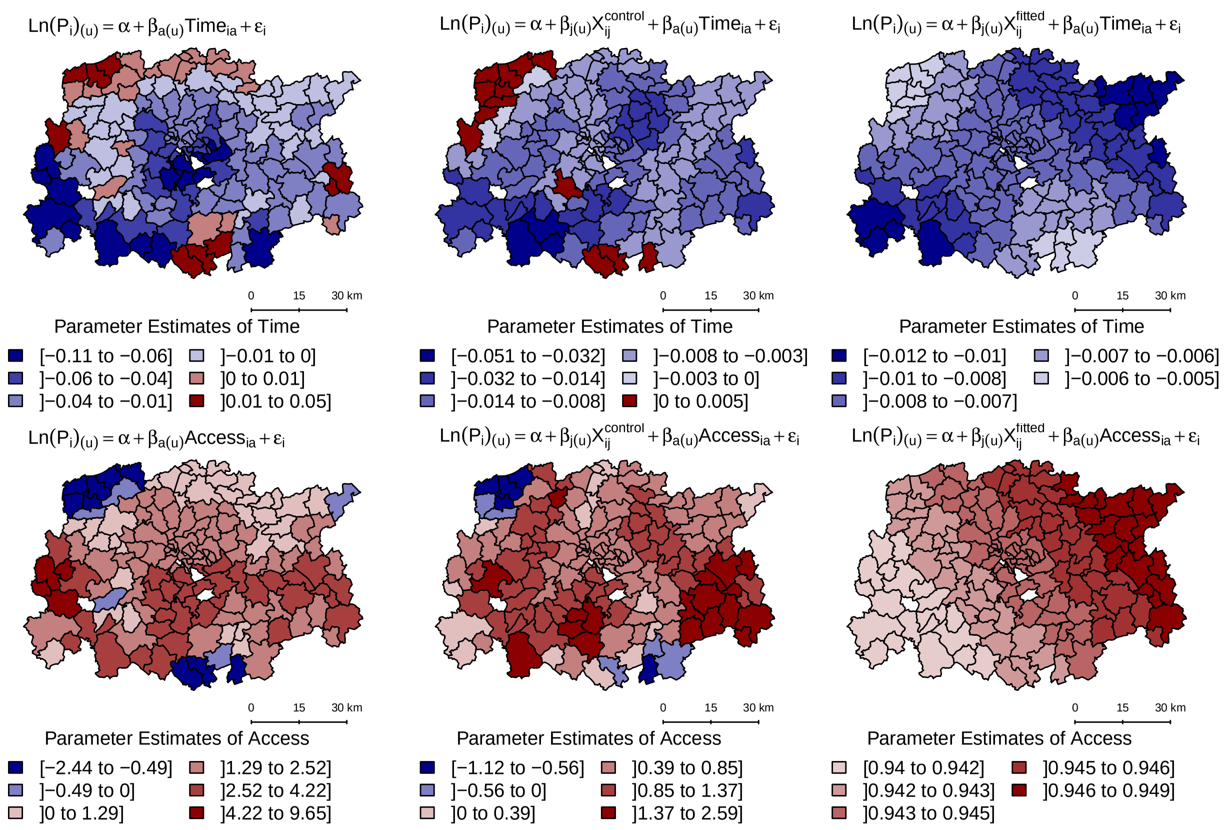

Figure 3.6 – GWR model (parameter estimates; top = travel time to the Brussels CBD; bottom = accessibility to jobs)

3.5 Discussion

3.5.1 Consistency and limitations

The Alonso-Muth model assumes that the utility of households' increases with the accessibility to jobs and decreases with land prices. The positive parameter estimates of the Accessibility to jobs are thus consistent. Since this variable takes into account all jobs in Belgium (see section 2.2), its significant influence on land prices for all delineations of Brussels was expected. No significant difference is observed in parameter estimates, and the predicted land price gradient is linear when expressed as function of this variable. These results are consistent with Ahlfeldt (2011), who found that gravity-based accessibility indicators are better at capturing the employment accessibility than the traditional distance to the CBD. Moreover, selection of the accessibility to jobs rather than the distance to the CBD by the stepwise procedure (case A, Table B.1) is also in line with the result obtained by Ahlfeldt (2011). Hence, on a general point of view, the Travel time to Brussels is less adequate to capture accessibility to jobs, since it only reflects the accessibility to the portion of the jobs located in the CBD.

This latter characteristic is, however, the reason why this variable is used to assess the efficiency by which the delineations of Brussels capture the influence area of that city. The results presented here are partially inconclusive on that aspect, since variations of the influence of accessibility to the CBD on land prices are weak. A possible explanation is the lack of detailed information on real estate transactions. As described in section 3.2.2, only the average value was available, and at the level of the municipalities (average size of 34 sq. km). Although Brussels is a fascinating case study, because of the numerous different delineations of the city proposed in the literature, data availability constraints reduce the robustness of statistical analysis. Note that this weakness is likely to influence the *UrbanSim* model of the city as well (see chapter 6) and raises the question of the adequacy of a disaggregated LUTI model to forecast the evolution of this city. This methodological choice, however, was constrained by the *SustainCity* project. Yet, even if parameter estimates do not show significant differences, the variables selected by the stepwise procedure and the slope of the predicted land price gradient can be used to compare delineations of Brussels.

3.5.2 Variations of the size of the study area

This chapter shows that no optimal delineation appears to estimate land prices. The adjusted R^2, and median error are similar for all delineations of Brussels larger than the Urban Region (Table B.2). Nevertheless, differences appear in the specification of the land price models. Delineations smaller than the Phone basin are limited to socio-economic or environmental amenities' factors, and the Travel time to the CBD does not have a significant influence. On the contrary, developable land and single-family houses prices are strongly influenced by Brussels in the entire province of Brabant (Vanneste et al., 2007). When including these municipalities in the study area, the Travel time to the Brussels CBD becomes a significant determinant of land prices (and it remains so for all larger delineations).

Therefore, the size of the study area influences the nature of the area included in the delineation of the city. Relatively small delineations of Brussels lead to a land prices' structure that corresponds to Hoyt's sector theory (1939). This structure can be explained by negative externalities (see Vandermotten et al., 2010). For large delineations, the land prices' structure is mostly consistent with the Alonso-Muth model, even if the accessibility to the secondary employment centre (for which the travel time to IC/IR stations is a proxy) has to be taken into account by the model.

3.5.3 Variations of the composition of the study area

The criterion used to define a delineation of Brussels influences the nature of the study area. Among large delineations of Brussels, delineations based on OD-matrices (Phone, Job, and MLA) have a steeper land price gradient than the RER zones (official and extended), Brabant and Union (Figure 3.5). The continuum of delineations shows an even lower slope of the land price gradient.

The latter group (RER zones, Brabant, and Union) has a larger proportion of their municipalities located in Flanders. All other things being equal, land prices tend to be higher than in Wallonia due to the scarcity of space and the denser urban network (Goffette-Nagot et al., 2011). Moreover, the complete Metropolitan Labour Area (MLA) of Leuven (12 municipalities) and Mechelen (5) are included in the Union of delineations of Brussels, together with eight municipalities belonging to the MLA of Antwerpen or Gent (see Figure B.2 in appendix). One municipality of the MLA of Charleroi (in Wallonia) is also included. For these municipalities, the percentage of people commuting (to work or to school) or migrating to the secondary centre is higher than to Brussels (Van Hecke et al., 2009), and no significant relationship is found between land prices and the Travel time to Brussels (Table 3.4).

The MLA of Brussels is also subject to some drawbacks. The threshold value used by Van Hecke et al. (2009), 15% of commuters to the Operational Agglomeration, is small compared to the one used in France (40%) or in the US (25%, see Dujardin et al., 2007). Hence, the Commuter Living Zone includes rural areas with relatively low land prices and population density (Brück, 2002). The relationship between land price and travel time to the Brussels to the CBD is thus weaker for that zone.

Zone	**Pearson correlation**		
	Travel time to Brussels	Accessibility to jobs	n
UR	-0.83***	0.85***	62
Union	-0.72***	0.85***	160
CLZ	-0.46***	0.71***	60
Other MLA	-0.27	0.63***	26

Table 3.4 – Variations of land prices with space (pearson product-moment correlation with the (log of) land prices; note: n = number of observations; *** = ρ significant at 0.001; CLZ = Commuter living zone of Brussels; Other MLA = Metropolitan Labour Area of Leuven, Mechelen, Antwerpen, Ghent or Charleroi)

3.6 Summary and implications for LUTI models

The chapter shows that delineation of cities exists for different purposes, and that they are not interchangeable. The methodology and criteria chosen to delineate a city have a clear influence on the size and the composition of the delineation produced and, as a consequence, on its nature (Table 3.5). It has long been noted that morphological delineations of cities (e.g. the Operational Agglomeration and the Fractal morphological agglomeration) do not capture the enlarged range of spatial interactions allowed by the modernisation of transport technology (Pumain, 2003).

Large delineations of cities can, however, be subject to an overestimation of the city's influence area. This latter problem arises from the inclusion of secondary centres and/or rural areas (see section 3.5.3). For instance, RER zones are based on the physical extension of the railway network, rather than on the actual commuting pattern. These delineations (and the continuum) imply that the influence of Brussels spreads evenly in all directions. Although this assumption holds for an isotropic and featureless landscape as the one of the Alonso-Muth model, this is clearly not the case in a highly urbanised country such as Belgium.

In terms of implications for operational applications of LUTI models, this chapter confirms the intuition that a poorly defined study area may affect the outputs of the model (see also chapter 5). The variations of parameter estimates observed between delineations of Brussels (Table B.2) mean that varying study areas will lead to different forecast of future real estate prices. These prices will, in turn, be used as inputs to forecast households and/or firms' location choices in *UrbanSim* (see chapter 5 and 6). Due to the sequential nature of *UrbanSim*, such biases are thus likely to propagate to main outputs of the model (i.e. the final location of the agents). Note that in *UrbanSim*, as in other LUTI models (e.g. MEPLAN), residuals of the real estate price sub model in t are inputed as independent variable in this sub model in $t+1$, therefore allowing

Criterion	**Size**	**Nature**
Morphological	Small	City centre, intra urban area
Socio-Economic	Medium	Urban region
Attractiveness (fluxes)	Large	Monocentric metropolitan area
Transport infrastructure	Large	Polycentric region

Table 3.5 – Nature of cities delineations

for a potential mitigation of this bias. Chapter 7 will discuss the optimal methodology to delineate the study area of an operational application of a LUTI model.

CHAPTER 4

Scale effect in a MNL model of employment' location choices

4.1 Introduction

Discrete Choice Models (or DCM) constitute a key component of most state-of-the-art LUTI models (see chapter 2). The reason is that they rely on the utility level perceived by agents and are, therefore, grounded in the micro economic theory. As in stand-alone applications (see Arauzo-Carod et al., 2010 for review), their purpose is to forecast location choices of agents (households or economic activities). Although this has been far less studied than for regression analysis (see chapter 2), they are subject to spatial bias when the choice set consists in areal units (which is precisely the case for LUTI models). The aim of this chapter is extending the state-of-the-art in that respect: how are DCMs influenced by a change of the size of the Basic Spatial Units (hereafter BSU) used as the choice set in a LUTI model context? To provide an answer as complete as possible to this question, this chapter considers three successive research questions.

First, do the parameter estimates of DCM vary with the size of the spatial units constituting the choice set? An empirical analysis of job's location choice is conducted for two spatial patterns (monocentric, using the jobs in services and polycentric, using industrial activities, see section 4.2). The study area is the metropolitan area of Brussels (Belgium), and four levels of hierarchical administrative units are used as BSU. For consistency purposes, the econometric framework is identical to the DCM implemented in *UrbanSim*, i.e. a linear-in-parameter, utility maximizing, multinomial logit model (see Waddell et al., 2003).

Secondly, are the variations in parameter estimates between BSU significant compared to misspecification issues? Amrhein (1995) and Briant et al. (2010), among others, show that this latter issue may be more important than the former one. Here, five different specifications are estimated for each location choice model. It allows a comparison of variations between BSU and specifications and, therefore, of the relative importance of spatial biases versus misspecification issues.

The third step is to assess operational implications. A clustering procedure is conducted to compare the structure of the probability of location through scales. Moreover, the following experience is performed: assuming that new jobs are created and are allocated through the BSUs proportionally to the predicted utility level, does the distribution of these jobs vary when simulated for different BSU level?

To cover these questions, the chapter is organized as follows: the case studies are detailed in section 4.2 and the methodology in section 4.3. Section 4.4 presents the results, which are discussed in section 4.5. Section 4.6 summarises the findings and their implications for LUTI models.

4.2 Case study

The city of Brussels (Belgium) is used as case study. Administrative and statistical delineations in Belgium have a high level of spatial detail (see section 4.2.1), allowing studying the effect of scale on a more continuous way than in previous works of Arauzo-Carod and Antolín-Manjón (2004) and de Palma et al. (2007).

Moreover, it can be expected that the sensitivity of DCM to scale will be affected by the underlying spatial distribution of firms/jobs: if jobs are concentrated in one main employment centre, it will always emerge from the neighbouring areas. On the contrary, if jobs are scattered through small employment centres, these secondary centres may be diluted within their neighbourhood for large BSUs. To assess this potential effect, two case studies are considered, corresponding to two different spatial patterns. The first case study (*Monocentric*) examines the location choices of jobs in the tertiary sector on a mono centric study area, the Urban Region of Brussels (see section 3.2). The combination of the high centrality of jobs in services and of a small study area results in a mono centric case, with most of the job concentrated in the CBD of Brussels (Figure 4.1a).

Case study	BSU level	n	Surface (km^2) Min	Mean	Max
Monocentric	Statistical ward	2 074	0.01	0.7	14.9
	Sections	550	0.01	2.6	15.4
	Former municipalities	173	0.25	8.7	45.0
	Municipalities	62	1.06	23.8	68.6
Polycentric	Statistical ward	4 223	0.01	0.9	15.9
	Sections	1 217	0.01	3.3	16.5
	Former municipalities	473	0.25	8.7	45.0
	Municipalities	126	1.06	32.4	96.4

Table 4.1 – Basic spatial units (n = number)

In the second case study (*Polycentric*) location choices of industrial jobs are estimated on a large and poly centric study area. The so-called RER zone (see section 3.2) has been used. Together with the less concentrated distribution of jobs observed for industrial activities than for services, the larger extent of this study area leads to a more poly centric structure (Figure 4.1b). Figure C.1 (in appendix) shows the extension of these studies areas, and the administrative units used as BSUs (see also Table 4.1).

4.2.1 Basic spatial units

Four levels of administrative and statistical units are used in the analyses. The higher level is the municipality. Each can be subdivided into "former" municipalities (aggregated into the current municipalities in 1977). For census purposes, those latter units can be divided into sections, themselves composed of several statistical wards. These are the smallest areal units for which data are available from the Belgian Directorate General Statistics and Economical Information. These BSU levels are hierarchical, meaning that a BSU of level n is strictly contained in only one BSU of level $n+1$ (see Figure C.1). Conversely, it means that statistical wards can be aggregated recursively into sections, former municipalities, and municipalities, without boundaries conflict. Since this chapter focuses on operational implications, we did not consider artificial territorial units (i.e. division of space based on raster, gridcells, or Thiessen polygons).

4.2.2 Jobs' location

For job's location data, the Home-To-Work Travel (HTWT) Survey of 2008 (see Witlox et al., 2011, Van Malderen et al., 2012) has been used. This survey is a legal requirement, which allows a response rate above 90%. For all firms located in Belgium having at least 100 employees, it gives the geographic coordinates of all plants of more than 30 employees. The NACE-BEL 2008 2-digits classification of economic activities (see DGSIE, 2011) has been used to select the firms in industrial activities (NACE code from 12 to 45 included) that compose the *Polycentric* case study, and in services (NACE codes higher than 45), for the *Monocentric* case study. Note that since the HTWT data set is limited to firms of more than 100 employees, it only accounts for 57% of the total number of jobs in tertiary sector and 43% for industrial activities (see ONSS, 2015). It should also be noted that the HTWT database includes all jobs at one given time, rather than jobs having recently relocated. Figure 4.1 shows the spatial distribution of jobs for the two case studies considered in this chapter. Descriptive statistics of the number of jobs per BSU are given by Table 4.2.

Note that few studies exist on the use of DCM to assess location choices of firms or jobs in Brussels, except Baudewyns (1999) and Baudewyns et al. (2000) who use a different framework (stated preferences). Nevertheless, Marissal et al. (2006) show that jobs remain concentrated in central places (see also Riguelle et al., 2007) even if (between 1991 and 2001) job's growth was systematically lower in the city centre than in the suburbs. Tertiary sector, i.e. the *Monocentric* case study (financial activities, in particular), is highly concentrated in the Brussels-Capital Region, while non-trade services are less concentrated but still reflect the distribution of the population and, consequently, the urban structure (Marissal et al., 2006). Secondary cities have a higher importance for industrial activities (i.e. the *Polycentric* case study), especially in Flanders.

4.2.3 Zonal characteristics

We would stress here that the goal of this chapter is not to find the best explanatory model for job's location choices in Brussels. The HTWT data set has been used because it was available, and it allows comparing different spatial patterns (mono centric and poly centric). Only simple variables are used as independent factors. Nevertheless, we attempted to rely on variables grounded in the economic geography literature. The econometric model used throughout this chapter (see section 4.3) follows the neoclassical perspective (Hayter, 1997), which assumes that agents are rational and have perfect information (see Shukla and Waddell, 1991, Waddell et al., 2003). In such conceptual framework, location determinants are cost-driving factors, i.e. agglomeration economies, transport infrastructure, and technology or human capital (see Arauzo-Carod et al., 2010 for review).

The zonal characteristics used throughout this work attempt to cover these three categories. A wide range of variables could be taken as a proxy, but two reasons explain those selected here. First, given the high variations in the sizes of the BSU (see Table 4.1), we restrained ourselves to independent factors expected to play a role at all scales. Local characteristics that could have been significant for small BSU alone (see e.g. de Palma et al., 2007) are thus excluded. Secondly, only variables that can be aggregated into large BSU by sums or means have been considered. It constitutes certainly one weakness of this work. In particular, one could wonder if the use of a more detailed model will not reduce variations across scales. This is, however, not our opinion, since previous work using more complex specifications found significant variations of parameter estimates between BSU (see e.g. Arauzo-Carod and Antolín-Manjón, 2004; de Palma et al., 2007). The use of a more advanced method is perhaps a better way, with the restriction that they are not, to the exception of nested logit, implemented in LUTI models, nor than in most operational applications of DCM (see chapter 7). The following paragraphs define the zonal characteristics used as independent factors into the location choice model.

For agglomeration economies, the density of jobs was selected. However, since a one time-step data set is used, and not firms having recently relocated, explaining the jobs' location by the jobs' density leads to major endogeneity concern. Preliminary analysis proved that including the job's density in the model precluded any other variables to have a significant effect. Hence, this variable has been excluded. Another problem is encountered for technology and human capital that mostly rely on socio-economic factors: the DGSIE only discloses real estate prices at the municipalities level. Most studies on real estate prices in Belgium (e.g. Goffette-Nagot et al., 2011, Cavailhès and Thomas, 2013, or chapter 3) thus use municipality as the level of analysis. There is no example of the estimation of a disaggregated indicator of real estate values at the statistical ward level (which will be a complete work in itself). Moreover, simply attributing to all lower-level BSU the value of the municipality to which it belongs may bias econometric estimations and do not seem a good option in a work dedicated to the *scale effect*. Hence, real estate prices will not be used in this work. Population density (POP_DENS), available from the DGSIE at the statistical ward level, is instead used as a proxy. It is defined as the number of inhabitants per square kilometre. Population density is likely to have a positive influence on utilities for large BSU (municipalities and former municipalities) since it will represent, at this scale, urban areas. However, for small BSUs, a negative influence can be assumed due to competition for land (a high population density meaning that there is no or few spaces left for other activities).

Transport and accessibility amenities are accounted for by four variables: travel time (TIME_BXL) to Brussels (in minutes), by car and congestion included is used as an accessibility indicator to the main employment centre. Travel times are computed between the centroid of each BSU and the centroid of the municipality of Brussels (data from Vandenbulcke et al., 2007). The accessibility to jobs (ACC_JOBS) is a Shimbel index of the travel time by car (data from Vandenbulcke et al., 2007) between i and all other spatial units of the same level in Belgium, weighted by the total number of jobs (self-employed excluded) located in these BSUs, from the HTWT database. Local amenities are accounted for by the Euclidean distance between the centroid of each BSU and (a) the closest IC/IR trains station[1] (DIST_TRAIN) and (b) to the nearest entry/exit on a highway (DIST_HGW). The use of the Euclidean distance is a simplifying assumption. Note that since most of the study area (for both *Monocentric* and *Polycentric* case studies) correspond to sub urban or rural areas, the use of the Manhattan distance would not have been a better option.

In Belgium, Baudewyns et al. (2000) found that the proximity of transport infrastructure has a positive impact on firms' location choice, and similar findings are numerous in the empirical literature (see Arauzo-Carod et al., 2010, for review). These variables are thus expected to have a positive parameter estimate for both *Monocentric* and *Polycentric* case studies and for all BSUs.

Figure 4.1 shows the distribution of these explanatory factors and their descriptive statistics are given by Table 4.2. For BSUs larger than statistical wards, the databases have been generated by aggregation of the initial data, by sums or means. Note that in the econometric model, all these variables are expressed in log.

4.3 Econometric estimations and sensitivity analyses

4.3.1 Location choice model

Fundamentals of DCMs are simple: an agent select one alternative among those available (the choice set), in order to maximize his utility at the time when the choice is made (Ben Akiva and Lerman, 1985). These alternatives have to be *mutually exclusive*, *exhaustive* and their number must be *finite* (Train, 2003), all conditions that hold for areal units. Our econometric framework is here identical to the Employment Location Choice Model in *UrbanSim* (see Waddell et al., 2003). It corresponds to the classical linear-in-parameters, utility maximizing MNL model (Ben Akiva and Lerman, 1985). No alternative specific constants are included. Due to the size of the choice set, a random sampling of 10 alternatives per observation is performed (the selected one, and nine non-chosen alternatives), as proposed by McFadden (1978) and, again, to mimic the specification implemented in *UrbanSim*. Hence, the probability that an

[1]A main train station is here defined as a train station where IC (intercity, fast direct trains) and IR (interregio, semi direct trains) train calls.

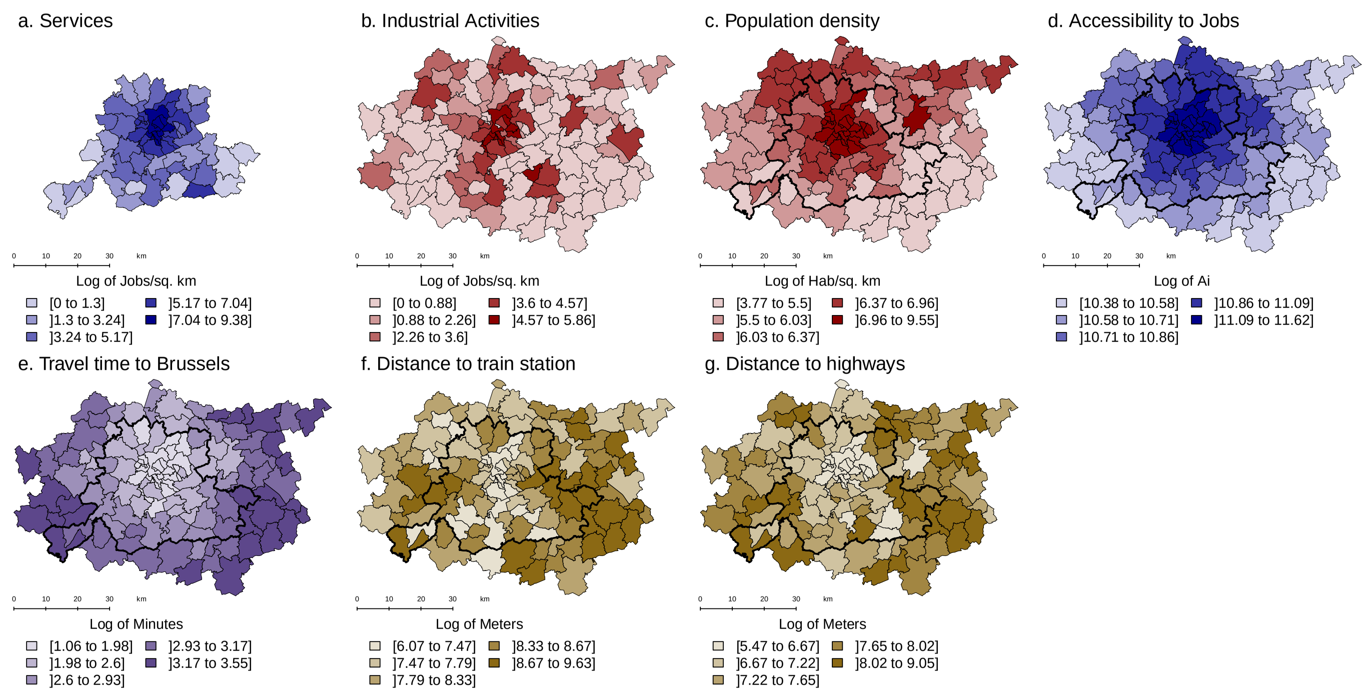

Figure 4.1 – Variables of the location choice model (note: municipalities level; all variables expressed in log; discretisation = quantiles)

BSU	Variables	Min	Mean	Max	SD
Statistical wards	SERVICES	0	165	10 620	679
(n = 2 074)	INDUSTRY	0	16	3 281	127
	POP_DENS	0.00	6.33	10.71	2.24
	INCOME	0.00	6.75	9.32	3.50
	LAND_PRICE	3.34	4.88	7.42	0.69
	TIME_BXL	0.00	2.68	3.80	0.81
	DIST_TRAIN	4.50	8.02	9.75	0.82
	DIST_HGW	3.57	7.34	9.49	0.98
Sections	SERVICES	0	622	24 238	1 887
	INDUSTRY	0	54	3 281	250
	POP_DENS	0.00	5.99	10.34	1.59
	INCOME	0.00	6.69	9.32	2.44
	LAND_PRICE	3.34	4.78	7.42	0.66
	TIME_BXL	0.01	2.78	3.80	0.66
	DIST_TRAIN	5.47	8.15	9.74	0.78
	DIST_HGW	4.40	7.50	9.48	0.94
Former Muni.	SERVICES	0	1 976	58 618	5481
	INDUSTRY	0	139	4 796	461
	POP_DENS	0.00	5.71	9.55	1.23
	INCOME	0.00	6.57	9.29	1.88
	LAND_PRICE	3.34	4.64	7.42	0.58
	TIME_BXL	0.00	2.92	3.80	0.54
	DIST_TRAIN	6.07	8.32	9.74	0.71
	DIST_HGW	5.21	7.72	9.48	0.83
Municipalities	SERVICES	0	5 515	103 675	13 629
	INDUSTRY	0	522	6 282	1 033
	POP_DENS	3.77	6.24	9.55	1.06
	INCOME	3.86	6.67	9.03	0.98
	LAND_PRICE	3.34	4.88	7.42	0.69
	TIME_BXL	1.06	2.71	3.73	0.57
	DIST_TRAIN	6.07	8.10	9.62	0.68
	DIST_HGW	5.47	7.45	9.30	0.81

Table 4.2 – Variables of the location choice model (by BSU levels; note: all variables expressed in log and values are given for the *Polycentric* case study; SD = standard deviation; SERVICES and INDUSTRY = number of jobs for, respectively, the *Monocentric* and *Polycentric* case studies)

alternative i is selected (P_i, see equation 4.1) depends on its utility (u_i), which itself relies on X_i, the characteristics of i.

$$P_i = \frac{e^{\beta_k X_{ik} + \epsilon_i}}{\sum_j e^{\beta_k X_{jk} + \epsilon_j}} \tag{4.1}$$

The model is based on an individual representation of jobs (rather than firms). Hence, no firm-specific factors are included and the only characteristics of the jobs taken into account are their current location. Explanatory variables are thus limited to site-specific factors. This choice matches those made in recent applications of the *UrbanSim* model, where the characteristics of the firm are not taken into account (see in particular Cabrita et al., 2015 for Brussels). It is, however, clear that the little effort made to consider jointly plant and zone factors remains one weakness of DCM (Arauzo-Carod et al., 2010; see e.g. Arauzo-Carod and Antolín-Manjón, 2004 for an analysis that consider both the size of the firms and of the areal units). Models for *Monocentric* and *Polycentric* case studies are estimated independently (in R, using the MLOGIT package; see Croissant, 2012).

Sensitivity analysis of parameter estimates

For each case study, the methodology was the following. Six different combinations of the independent variables have been drawn (Table 4.3). They focus on socio-economic characteristics, on accessibility indicators or on a mix of these factors. Two reasons explain the use estimation of different specifications. First, Amrhein (1995) and Briant et al. (2010) argue that for econometric model the misspecification's issue induces larger variations of parameter estimates than those observed between BSUs. It was thus necessary to test this issue here (which corresponds to our second research question). Secondly, no study on jobs' location choices based on a DCM model exists for Brussels (see section 4.2.2).

In the subsequent analysis, all independent factors are expressed in log. This choice has been made since the goodness-of-fit of the model was generally higher than with the linear form. However, we did not perform a formal test on the functional form of the model. The relative influence of this latter issue on parameter estimates, compared with the scale effect or mis specifications issue remains thus an open question.

These specifications are estimated for the four levels of BSUs. The benchmark model (i.e. the one used in the sensitivity analysis of the DCM to the size of the BSUs, first research question) is selected among the estimated specifications using the following conditions. (a) The AIC has to be lower, or similar to the other specifications. The AIC is used here thanks to its ability to compare the goodness-of-fit of specifications involving different number of independent

	Specifications				
Variables	(1)	(2)	(3)	(4)	(5)
POP_DENS	0	1	0	1	0
ACC_JOBS	1	1	0	0	1
TIME_BXL	0	0	1	1	1
DIST_TRAIN	1	1	1	1	1
DIST_HGW	1	1	1	1	1

Table 4.3 – Specifications of the location choice models (1 means that the variable is included in the specification, 0 otherwise)

variables. (b) All independent variables should have a significant effect on the utility level (for $\alpha = 0.05$). Other specifications (hereafter referred as control) will be used to compare the magnitude of the variations of parameter estimates between BSUs to those observed between specifications (second research question). Both the direction and magnitude of these variations are examined. Direction consists in studying whether the parameter estimates increase or decrease with the size of the BSU and if change of signs can be observed (between BSUs and between specifications). The magnitude refers to the absolute differences between parameter estimates. In particular, we aim to identify which pairs of parameter estimates are significantly different from each other, between BSUs and between specifications, by pair wise t-tests (Bonferroni correction of the p-values).

Sensitivity analysis of the probability of location

The last step is to assess operational implications (third research question). In LUTI models using DCM to forecast location choices of jobs, the predicted probabilities of location (equation 4.1) are used to distribute new and/or relocating jobs among the BSUs (Waddell, 2002; Waddell et al., 2003). On a pure operational point of view, it can thus be argued that the variations of parameter estimates through scales are of little importance as long as the spatial structure of these predicted probabilities of location remains identical. Let's imagine a municipality composed of 10 statistical wards. If the sum of the predicted probability of location at this level is equal to the probability predicted for the municipality, the scale does not influence their spatial structure.

Moreover, the sum over all alternatives of the individual probability of location is equal to one (equation 4.1). An increase in the utility of one zone will thus (all other things being equal) increase the probability of that zone and decrease those of all other zones. Hence, the link (through utilities) between variations of parameter estimates and predicted probability of location is not a direct one. Let us add that multivariate specifications are used, meaning that an increase of a parameter estimate can be compensated by a decrease of another one. Descriptive statistics of variables also change between the four levels of BSU. These reasons make it difficult to identify the exact influence of the variations of parameter estimates. Operational implications of the choice of the BSUs on LUTI models using DCM to forecast job's location choices are thus assessed using the predicted probability of location, by a two-step procedure.

First, a cluster analysis (Ward method - note that other hierarchical clustering methods have been tested, leading to similar results) has been realised. The observations used are the statistical wards, each being characterized by its probability of location (predicted by the benchmark model) and by the probabilities of location of the three larger BSU to which it belongs. It has the advantage of allowing for a finer spatial level of analysis. Another benefit is that it allows taking into account the four levels of BSU, rather than conducting two-by-two comparisons. The optimal number of clusters is determined by the combination of CCC (Sarle, 1983), pseudo-t^2 (Duda and Hart, 1973) and CH index (Calinski and Harabasz, 1974). The underlying idea is that a similar spatial structure of the predicted probability of location should lead to a linear progression of descriptive statistics per cluster. That is to say, that one cluster should have relatively low probability of location for all BSU levels, another medium probability, and so on. A cluster corresponding, for instance, to statistical wards having a low probability of location for small BSUs but a high one for large BSUs means, on the contrary, that the spatial structure of potential employment centres varies through scales.

Secondly, the following exercise is conducted: an increase of 1% of the number of jobs is assumed (because of economic growth), and these new jobs are randomly distributed among BSUs, each BSU being weighted by its probability of location predicted by the DCM. Again, this procedure mimics those employed in LUTI models (see Waddell et al., 2003). The predicted number of new jobs per municipality can then be compared to the one per statistical ward, by aggregating the latter one at the municipality level. To mitigate the stochastic variations, 100 repetitions of the distribution procedure are used. Note that relocation are not allowed here, meaning that the location of existing jobs is fixed.

The workflow of the sensitivity analysis can be summarized as follows: (1) draw of a set of specifications, (2) estimation for the four levels of BSU, (3) selection of the benchmark model, (4) analysis of the parameter estimates' variations through scales, (5) analysis of the parameter estimates' variations across specifications, (6) cluster analysis, and (7) experience of the distribution of new jobs. The estimations have been repeated over 100 independent samples of 1% of the observations (in the further analysis, the mean parameter estimate over the 100 samples is used).

4.4 Results

4.4.1 Selection of the benchmark model

Figure 4.2 gives the AIC values for each specification. For the *Monocentric* case study, the variations observed across specifications are never significant. Non-significant parameter estimates (see Table C.1) are found for specifications (2), (4), and (5). Moreover, specifications (1) and (5) exhibits a multicollinearity problem between the accessibility to jobs and either the distance to highways or the travel time to Brussels, leading to negative parameter estimates for the former variable. Hence, specification (3) will be used as a benchmark for further analysis of the variations through scales, since (a) its goodness-of-fit is similar to other specifications, (b) all the parameter estimates are significant, and (c) of the expected sign (the utility decreases when the distance to Brussels or to transport infrastructure increase).

For the *Polycentric* case study, at the statistical ward level, the AIC value is significantly lower for specifications (2) and (4). Non-significant parameter estimates are found for all cases (see Table C.2), but less frequently for specifications (2) and (3). The multicollinearity issue remains present, but its magnitude is reduced. Hence, for comparability purpose with the *Monocentric* case study, it has been decided to use specification (3) as a benchmark also here. Other specifications are used as control, to compare variations linked to the size of the BSU with the variations between specifications.

Note that the McFadden pseudo-R^2 (see Table C.1 and C.2) of the benchmark model is in most cases slightly lower than the one of specifications (2) and (4), but the differences remain weak, especially since these specifications include 4 independent factors instead for three for the benchmark.

4.4.2 Variations of estimate parameters

Between BSU levels

Across BSUs, parameter estimates of the benchmark model are significantly different (at the 5% level) for all variables, on all pairs of BSUs and for both *Monocentric* and *Polycentric* cases. The only exception is the Statistical wards/Sections pair for DIST_HGW (results of the pair wise *t*-test comparisons are given in Table C.3, in Appendix).

Parameter estimates of the benchmark model do not evolve monotonously with the size of the BSUs. For the *Monocentric* case, Municipalities appear to behave differently than the three smaller BSUs, especially for DIST_TRAIN and DIST_HGW (Table C.1). For the *Polycentric* case (Table C.2) depending on the variable, Statistical wards and Municipalities appear different from the other BSUs. No changes of sign are observed among parameter estimates of the benchmark model. It should be noted, however, that TIME_BXL evolves

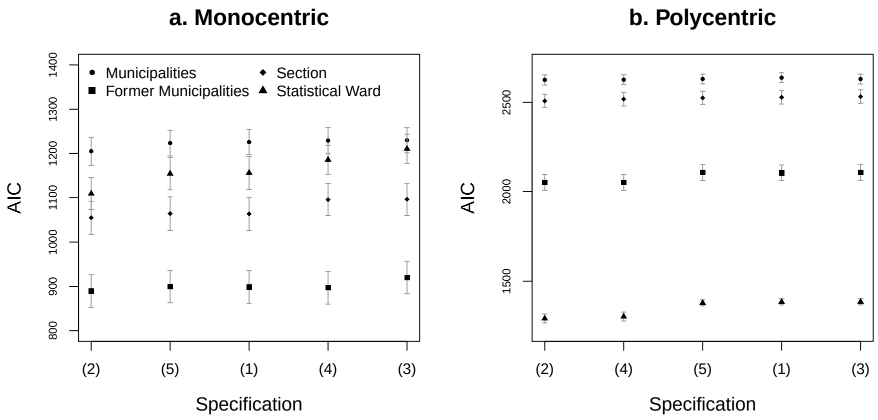

Figure 4.2 – **Variations of the AIC** (note: error bars = AIC value +/- one standard deviation; (x) refers to the specification, see table 4.3)

from a non-significant to a positive effect with the size of the BSUs for the *Polycentric* case study.

Between specifications

Most parameter estimates are also significantly different between the benchmark model and control specifications for most variables and BSUs (Table C.4, in Appendix). Non-significant differences appear, however, more frequently, but no clear explanations for their variations across case studies, specifications, or BSU levels emerge.

Changes of sign among parameter estimates remain limited, for the *Monocentric* case study, to the Former municipalities' level: the parameter estimates of TIME_BXL are positive for model (5), but negative in the benchmark model. The same opposition can be observed for the DIST_TRAIN variable between model (1) and the benchmark. For the *Polycentric* case, no change of signs can be observed for TIME_BXL and DIST_HGW, only evolutions from significant to non-significant. The DIST_TRAIN variable shows opposite parameter estimates between model (2) and the benchmark at the Former municipalities and Municipalities level. As one could have expected, differences in parameter estimates appear to be linked to the degree of similarity between specifications in terms of variables included (such as the benchmark model and model 4). The inclusion of only one additional variable may, however, have a high influence on parameter estimates, as showed by the pair of specifications (1) and (2), and between the benchmark model and model (5).

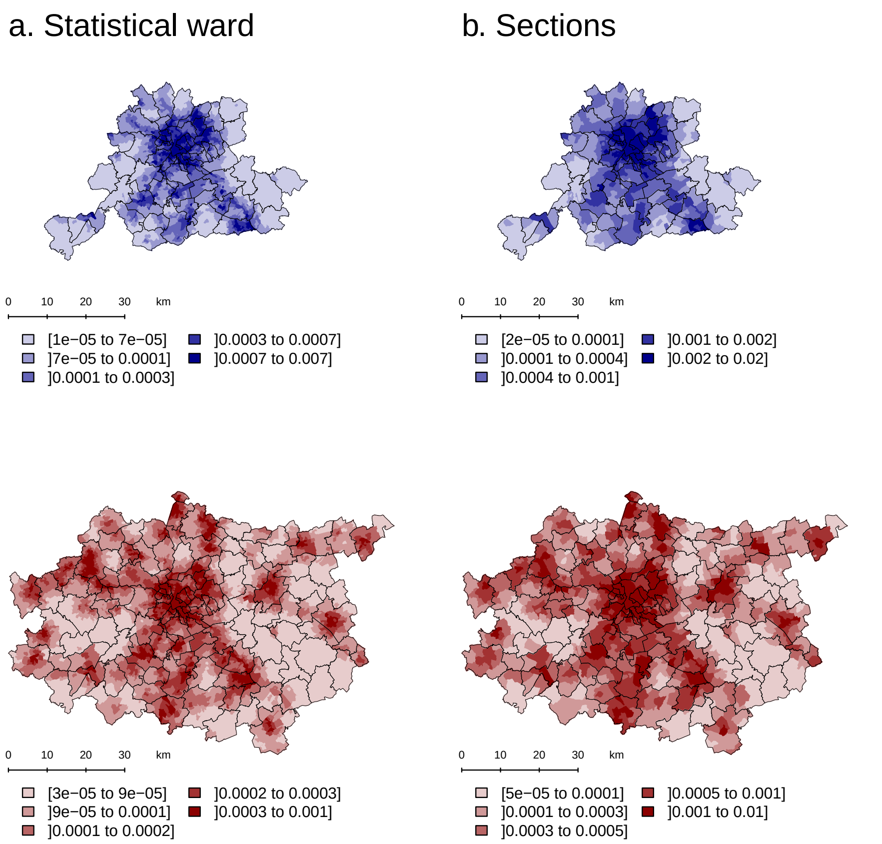

Figure 4.3 – Predicted probability of location (top = *Monocentric* case study: bottom = *Polycentric*; discretisation = quantile)

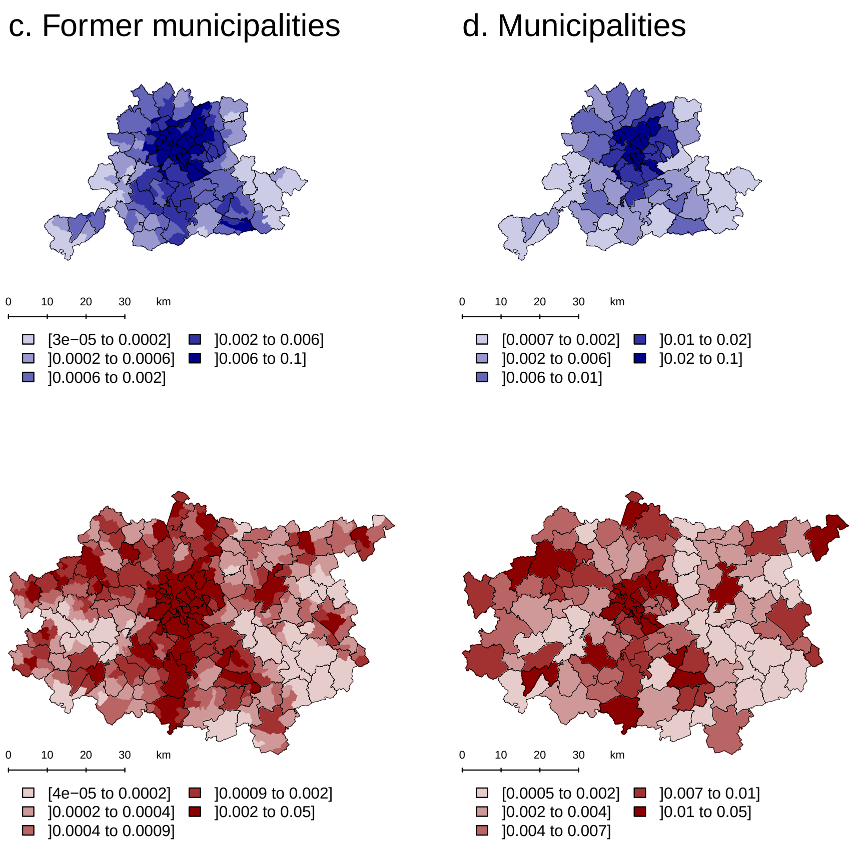

Figure 4.4 – Predicted probability of location (continued from previous figure; top = *Monocentric* case study: bottom = *Polycentric*; discretisation = quantile)

4.4.3 Predicted probability of location

Using the benchmark, the highest probabilities of location are found in the Brussels city centre, which is consistent with the sign of parameter estimates (Table C.1 and C.2). Other clusters of BSUs with high probabilities of location can be observed close to train stations and/or highways (the combination of both factors corresponding usually to a secondary city). Variations in the spatial structure of the probability of location through BSUs levels can also be observed on the maps of the predicted probability of location (Figure 4.3).

Note that the aggregation of the predicted probability of location from statistical ward to municipalities' level (by summing the probability of all statistical wards belonging to the same municipality) results in values different than those directly predicted at municipalities' level. Relative differences vary from -336 to +88% for the *Monocentric* case study (mean = -21%), and from -324 to +86% for the *Polycentric* one (mean = -18%). Moreover, the correlation (Pearson) between aggregated and direct values is medium: 0.61*** for *Monocentric* and 0.52*** for *Polycentric*. Hence, to explore these variations on a consistent way for the entire study area, a clustering procedure has been conducted.

For the *Monocentric* case study, three clusters are obtained. They are organized in concentric rings around the centre of Brussels (Figure 4.5) and correspond respectively to relatively low ($CL1_m$), medium ($CL2_m$), and high ($CL3_m$) probabilities of location (see Figure C.2, in Appendix). Probabilities are significantly weaker in $CL1_m$ than in $CL2_m$, and in $CL2_m$ compared to $CL3_m$, for all BSUs (at $\alpha = 0.05$). For the *Polycentric* case study, the clustering produces five clusters. The concentric structure from high to low probabilities also appears, with $CL3_p$ being the city centre of Brussels, $CL1_p$ rural areas, and $CL2_p$ suburbs or secondary centres (Figure 4.5). Two particularities should, however, be noted: $CL4_p$ and $CL5_p$ have similar values for small BSUs, but relatively low values are observed at the municipalities level for $CL4_p$, and the opposite for $CL5_p$ (Figure C.3). Note that for other specifications, the number of clusters (using the exact same procedure) varies from 4 (model 2 and 5) to 10 (model 4) for the *Monocentric* case study, and from 4 (model 5) to 11 (model 4) for the *Polycentric* one. The spatial pattern is also similar, although variations in the number of clusters make a formal comparison difficult.

Figure 4.6 shows the differences in the predicted number of new jobs between Municipalities and Statistical wards. Negative differences mean that more new jobs are predicted at the Statistical wards' level than at the Municipalities' level (and positive differences the opposite). The spatial structure of the variations is similar for the two activity sectors, which was expected since (a) identical specifications are used and (b) the parameter estimates are of the same sign. The correlation (Pearson product-moment) between the number of new jobs predicted by Statistical wards and Municipalities is of 0.62*** for the *Monocentric* case study, and also of 0.62*** for the *Polycentric* case study. Greater

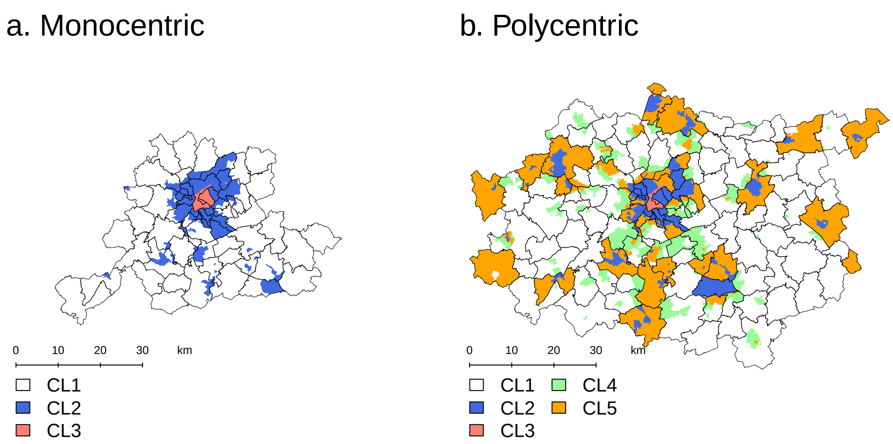

Figure 4.5 – Clustering procedure (at the statistical ward level, using the predicted probabilities of location)

absolute variations are found for the former one. They can be explained by the larger number of new jobs (3 419 versus 658) and by the lower number of BSU (62 versus 126). The municipality of Brussels and secondary cities receives fewer jobs at the municipality level than at the statistical ward level. On the contrary, more jobs are distributed (at the Municipalities level) in different small municipalities within the Brussels-Capital Region. In suburban or rural areas, the differences are limited in magnitude. Negative differences are found for municipalities close to transportation infrastructure, and positive differences for more peripheral municipalities. Hence, for our benchmark model, the concentration of jobs in cities appears to decrease with the size of the BSUs.

4.5 Discussion

4.5.1 Consistency and limitations

The sensitivity analysis presented in this chapter suffers from several shortcomings. The literature shows that the econometric framework used (linear-in-parameter MNL model) is subject to many limitations when applied on spatial choice sets (see chapter 7). It has specifically been decided to stick to this model, since it is the one used by many LUTI models. Nevertheless, one may wonder if the best option would not be to implement in LUTI model's specifications allowing to take into account a spatial effect (see e.g. Guo and Bhat, 2004; Guo and Bhat, 2007; Sener et al., 2011; Alamá-Sabater et al., 2011;

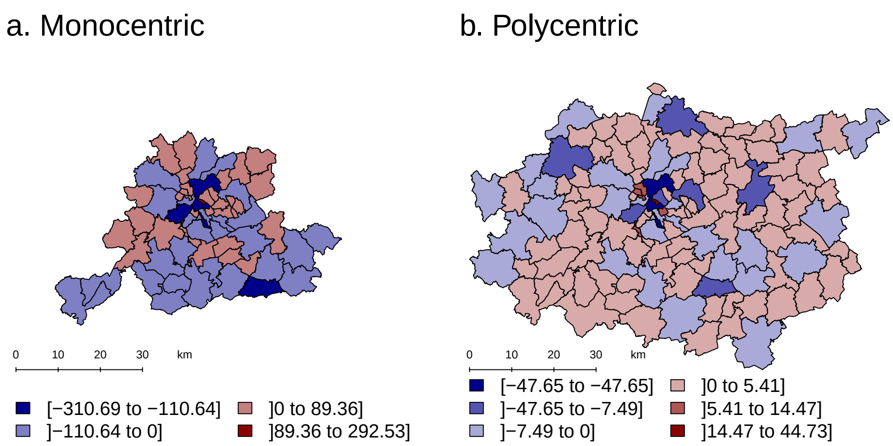

Figure 4.6 – Variations in the number of new jobs (Municipalities minus Statistical ward; discretisation = Jenks)

see also chapter 7), or nested logit models (Cornelis et al., 2012). Another drawback is that the specification of the model is limited to simple variables and is identical for all BSUs, while Arauzo-Carod and Antolín-Manjón (2004), de Palma et al. (2007), and our own findings, show that location choice factors do not act uniformly through scales.

The choice of keeping the same specification for all BSUs comes from the fact that the benchmark model performs better at all scales than any other estimated specifications. However, it can be argued that specifications tailored for each BSU could reduce the variations in the spatial structure of the probability of location. More fundamentally, it illustrates the limitations of the econometric specification currently implemented in *UrbanSim* (see chapter 6).

The variations observed among parameter estimates between BSUs were expected, given previous works on the Modifiable Areal Unit Problem (Amrhein, 1995; Arbia, 1989; Fotheringham and Wong, 1991; see, in particular, Arauzo-Carod and Antolín-Manjón, 2004 and de Palma et al., 2007 for DCM). This chapter thus confirms that such variations can be expected in all applications of DCM. The size of the BSUs does not influence the sign of parameter estimates here, meaning that the influence of a given factor on utility remains positive or negative through scales even if its intensity varies. However, independent factors were limited in this chapter to variables expected to have an influence on jobs' location choices at all scales. This result may thus be an artefacta due to the methodological choices made throughout this chapter.

4.5.2 Variations across BSU levels and specifications

The exact direction and magnitude of the observed variations in parameter estimates are obviously specific to our case study. It should be noted, however, that these variations are not monotonous. For the *Monocentric* case study, parameter estimates generally increase (in absolute terms) from Statistical wards to Former municipalities, while they decrease for Municipalities. For *Polycentric*, an increase is observed from Statistical wards to Former municipalities, followed by stabilization.

Other works assessing the influence of the MAUP on DCM (Arauzo-Carod and Antolín-Manjón, 2004 and de Palma et al., 2007) rely indeed on only two different levels of analysis. Hence, predicting or controlling variations of parameter estimates through scale is not straightforward, since these variations do not seem to be directly linked to the size of the BSUs. A potential reason is that administrative units do not correspond to the land use structure, even if modellers are often constrained to use such areal units, for a data availability reason, and because they remain a relevant unit for policy making. In the absence of comparable analysis in other works, it is difficult to assess the extension of these findings. Since the probability of location in a zone i depends on the utility of i and that of all other zones (see equation 4.1), we expect that they will remain valid for other case studies.

Previous works (Amrhein, 1995; Briant et al., 2010) found larger variations of parameter estimates between specifications than between BSU levels. Magnitude of the parameter estimates' variations are here comparable between BSUs and specifications (significant differences are observed in most cases; see Tables C.3 and C.4). The use of a different econometric model may constitute an explanation, and scales are not perfectly comparable. Briant et al. (2010), for instance, worked on France, with larger areal units. The high degree of similarity between specifications in terms of independent variables is also likely to reduce the differences between parameter estimates. Hence, spatial biases are here found to be of comparable magnitude with misspecification issue. The location choice models remain, however, very simple. The comparison of different specification should thus be seen as a methodological precaution (to be sure that spatial biases are worth worrying about), and certainly not as a complete analysis of misspecifications issues in DCM.

4.5.3 Spatial structure of the probability of location

The third research question addressed by this chapter is to examine if the size of the BSU influences the spatial structure of the predicted probability of location (let us recall that the case studies correspond to two activity sectors, services and industrial activities). The answer appears to be yes. More precisely, aggregation into large BSU levels leads either to a dilution or to a diffusion of the high probabilities predicted for small BSU levels.

Consider the example of a large BSU made of five smaller BSUs, and imagine that the estimation of the DCM at the small BSUs level leads to one BSU having a high probability of location, the four others a low probability. It is likely that at the aggregated BSU level, we will end up with a low probability of location. This is the "diffusion" process, that occurs when statistical wards with a high probability of location are rare in a given municipality. The "diffusion" process takes place when the importance of these statistical wards is larger (e.g. when two or three of the small BSUs, among the five, have a high probability of location). In that situation, these high probabilities are transferred to the aggregated BSU level, ending up with a municipality having a relatively high probability of location.

Results show that these processes have a limited influence for the *Monocentric* case study, where one cluster corresponds to relatively low probabilities of location at all scales, another to medium probabilities, and the third one to high probabilities. Moreover, the concentric structure of these clusters, centred on the Brussels CBD, is consistent with the urban structure. Statistical wards belonging to the "medium" cluster ($CL2_m$) that are scattered within the "low" one ($CL1_m$) encompass secondary employment centres (Wavre, Louvain-La-Neuve and Halle). Note that the distribution of the predicted probability of location is highly skewed, which explains why $CL2_m$ shows a negative deviation from the global median (Figure C.2, in Appendix).

The situation is more complex for the *Polycentric* case study (i.e. jobs belonging to industrial activities). "Low" to "High" clusters are also found (respectively $CL1_p$, $CL2_p$ and $CL3_p$) and their spatial extension are close to the one observed for the *Monocentric* case. This similarity can be explained by the fact that identical specifications are used for both case studies, and that the parameter estimates have the same sign. The study area being larger, the "Medium" ($CL2_p$) cluster encompasses extra secondary cities: Aalst, Mechelen and Leuven, while the "Low" ($CL1_p$) cluster corresponds to most of the rural parts of the study area.

Two additional clusters are observed. $CL5_p$ corresponds to statistical wards for which the probability of location tends to increase with the size of the BSU, and $CL4_p$ to the opposite. Looking at their spatial structure, $CL5_p$ is composed of low-density statistical wards located within the municipality of above mentioned secondary employment centres. (The statistical wards where the jobs are actually located in these municipalities belonging to $CL2_p$). $CL4_p$ is found surrounding isolated $CL2_p$'s statistical wards (or next to $CL1_p$), but located on the other side of a municipality boundary. Hence, the relative extension of potential employment centres depends on the scale of the analysis. On the first hand, when the size of the BSUs increases, some statistical wards could become part of an employment centre: this is the diffusion process observed for $CL5_p$. On the other hand, some statistical wards are diluted into a rural neighbourhood, as for those belonging to $CL4_p$.

4.6 Summary and implications for LUTI models

Operational implications, for LUTI models, of the sensitivity of DCM to the size of the BSU can only be partially explored by the stand-alone DCM presented in this chapter. The main reason is that the utility of each BSU is assumed here to be constant. There is no feedback's effect decreasing the utility of one BSU when its number of jobs' increases. In a complete LUTI model, feedbacks can arise from several factors (see chapter 5 and chapter 6).

A classical example is that an increase of the job density in one BSU should increases the real estate prices. If these real estate prices are included as independent variable, with a negative parameter estimate, in the specification of the location choice model, the utility level of that BSU will decrease in $t + 1$. Another potential feedback is that the travel time to Brussels may increase when the number of jobs increases, due to congestion effects simulated by the transport component of the LUTI model. If such feedbacks are present, the distribution of new jobs by a LUTI model would not be identical to the one simulated here.

The distribution of new jobs among BSUs, proportionally to the predicted utility level, vary when observed at different scales. For our case studies, substantial differences are observed in the absolute number of new jobs per municipality. A strong spatial structure also emerges, large BSU levels leading to a lower concentration of jobs in urban areas (Figure 4.3). Hence, even if the experiment performed here shows that the distribution of new jobs is similar through scales (high correlation for the number of new jobs per BSU, see section 4.4), it also shows that employment centres gain more or less importance during the simulation, depending upon the size of the BSU used. The nature of the BSUs for which the DCM predicts a high probability of job's location can explain these findings. Such BSUs correspond either to (a) actual employment centres (i.e. BSUs where a large number of jobs are located) or (b) to BSUs having similar intrinsic characteristics than these employment centres, even if the number of jobs located in it is presently limited. The latter ones are less frequent for large BSU levels than for small ones, for two reasons. First, a larger size means that adjacent BSUs are more likely to be dissimilar. Secondly, the lower number of large BSU means that the distribution of the probability of location is less continuous. Hence, the spatial heterogeneity increases with the size of the BSUs (although the variation range of independent factors is lower, see Table 4.2), which may explain the higher heterogeneity in terms of predicted probability of location. However, this spatial heterogeneity does not explain the differences observed when the probability of location at the statistical wards' level is aggregated into municipalities (see section 4.4.3).

LUTI models such as *UrbanSim* (see chapter 2) rely on the probabilities of location estimated by the DCM to distribute new/relocating agents during iterations of the model. Variations in the spatial structure of the high probability of location have thus important operational implications for land-use planning: they mean that estimating a jobs' location choice model for one BSU level instead of another may lead to forecasting different zones as the bet potential for future jobs' location. These variations in the spatial structure of the probability of location are not straightforward to predict, as showed by the structure of the clusters for the *Polycentric* case study. The reason is that they depend simultaneously on three elements: (1) the variations of parameter estimates over scales and (2) of descriptive statistics of the explanative factors, that affect the utility level of each BSU, and (3) the number of these BSUs: the sum of the probability of location is one. Hence, all other things being equal, an increase of the utility of one BSU leads mechanically to a decrease in all others.

Although the lower importance of the employment centre observed for large BSUs is a result specific to our case studies, this situation corresponds to the identification of employment sub centres. Hence, prior to estimate a DCM of jobs or firms' location choices, a careful exploratory spatial data analysis of the distribution of jobs should be conducted, in order to identify these sub centres at different scales and to compare their importance and localisation. Even if economic activities still tend to cluster into office parks (Archer and Smith, 1993), a multi polarisation trend has long been observed in cities (see e.g. Ladd and Wheaton, 1991), and many studies have attempted to identify the sub centres of employment. Nevertheless, no consensus appears on the appropriate methodology (Redfearn, 2007): traditional cut-off approach such as in the seminal work of Giuliano and Small (1991) on Los Angeles, locally weighted regressions (McMillen, 2001; McMillen and Smith, 2003), local measure of spatial autocorrelation (LISA, see Anselin, 1995; Riguelle et al., 2007) or Discrete Choice Model (Shukla and Waddell, 1991 with Dallas-Forth Worth as the case study). Given the sensitivity of econometric method parameter estimates to the size of the areal units demonstrated in the literature, the use of non-parametric methods (such as the LISA) should be preferred.

The sensitivity of Discrete Choice Models to the size of the spatial units used as the choice set highlighted by this chapter is consistent with the literature on the Modifiable Areal Unit Problem (see e.g. Arbia, 1989; Fotheringham and Wong, 1991) and can be summarised as follows. First, a significant influence to the size of the BSU is found for parameter estimates of the DCM and, consequently, on predicted probability of selection of the alternatives in the choice set. It allows extending previous works (Arauzo-Carod and Antolín-Manjón, 2004; de Palma et al., 2007; see also chapter 2) to a broader range of scales, by showing that similar conclusion can be draw from an urban case study with small BSU. Secondly, these variations are of the same order of magnitude than those observed between specifications. If we compare these results to those of Amrhein (1995) or Briant et al. (2010), it suggests that the relative importance of spatial biases and misspecifications issues depends on the case study and econometric methods considered. Here, a comparable influence on the model is found. Finally, the distribution of new jobs among the study area (using the probability of location predicted by the DCM) is different between scales. Since DCMs are used to forecast agent's location choices in many LUTI models (see Wegener, 2004 and chapter 2), their outputs (e.g. the final number of jobs and households per BSU) may thus be affected by using one level of BSU instead of another, which will be assessed in chapter 6.

Part III

Sensitivity of LUTI models' outputs

CHAPTER 5

Experiments on a synthetic case study

5.1 Introduction

This chapter focuses on the sensitivity to spatial bias of the direct outputs of a LUTI model, i.e. the final number of agents per zone. Chapters 3 and 4 show that such bias affect parameter estimates of econometric sub models within *UrbanSim'*. Chapter 4, in particular, demonstrates that the size of the areal units used as choice set may affect the estimation of agents' location choices. However, due to the large number of feedback effects within a LUTI model, it is not straightforward to link the variations of parameter estimates to those observed in the final distribution of agents.

A simple synthetic case study is used, in two configurations. First, a mono centric and isolated city. The influence of both the *scale* and *boundary effects* will be assessed, as well as the influence of two simple scenarios (improvements of the transportation network and urban growth boundaries). Secondly, a poly centric metropolitan area will be implemented. The idea of the second configuration is to study more in-depth the influence of the *boundary effect*.

The chapter is organised as follows. Section 5.2 details the internal principles of the *UrbanSim* model. The implementation of the synthetic city is presented in section 5.3, on both theoretical and practical point of view. Section 5.4 then presents the experiments on the mono centric configuration, followed by section 5.5 for the poly centric one. Section 5.6 summarises the findings and their implications for real world applications of LUTI models.

5.2 The *UrbanSim* model

UrbanSim is a quasi dynamic model, based on a disaggregated representation of agents (households, jobs, see Waddell et al., 2003; Simmonds et al., 2013) and space (buildings, zones). An overview of the model system (Figure D.6) can be found in Waddell (2000); Waddell (2002); or Waddell (2011). Figure 5.1 details the sequence of sub models called by *UrbanSim*. See chapter 2 for a comparison of *UrbanSim* with other LUTI models. Technical details on the sub models can be found in Sevcikova et al. (2007). Finally, note that *UrbanSim* only models the evolution of agents and land-use and has thus to be interfaced with a transport model, here *MATsim* (see Nagel et al., 2008), for simulations of the home to work trips and computation of accessibility indicators.

The model exists in three different versions: "gridcells" (the original implementation, discarded in recent applications), "zones", and "parcels". In the gridcells version, the study area is divided into pixels, while in the zone version, the BSUs are of irregular sizes and shapes, corresponding usually to administrative units. The parcel version is very similar, except for the real estate development sub model (see Zollig Renner and Axhausen, 2015). In this section, we focus on the zone version, since it is the one used in most applications of the model. Appendix D.1 gives the list of the tables required by *UrbanSim*' database (and their content), while the list of existing sub models is given in appendix D.2.

Hierarchical structure: zones, buildings, agents

UrbanSim operates on three levels: zones, buildings, and agents. The **zones** are the BSU used by the model, corresponding usually to administrative delineations. Their size and shape are, therefore, constant. These zones are used to store environmental amenities that consist in user-defined variables plus the outputs of the transport model (see appendix D.1). The surface of a zone is not explicitly shared among land use categories. *UrbanSim* relies rather on intermediate entities (between zones and agents), called buildings.

These **buildings** are divided among several building types (whose categories are user-defined). A building of type j in a zone i constitutes an aggregated representation of all individual buildings of that type existing in the zone. The buildings represent current and potential land use through different characteristics. Let us give an example assuming that a zone i comprises 10 houses (among other buildings) and that a provision (determined by land-use planning rules) exists for two new construction. The corresponding "house" building in i is characterised by the existing number of residential units (here 10, whether they are actually occupied or not) and by a capacity of residential units of 12. Other characteristics include the total land area consumed by these 12 houses, and their average value (see appendix D.1). Additional characteristics can be

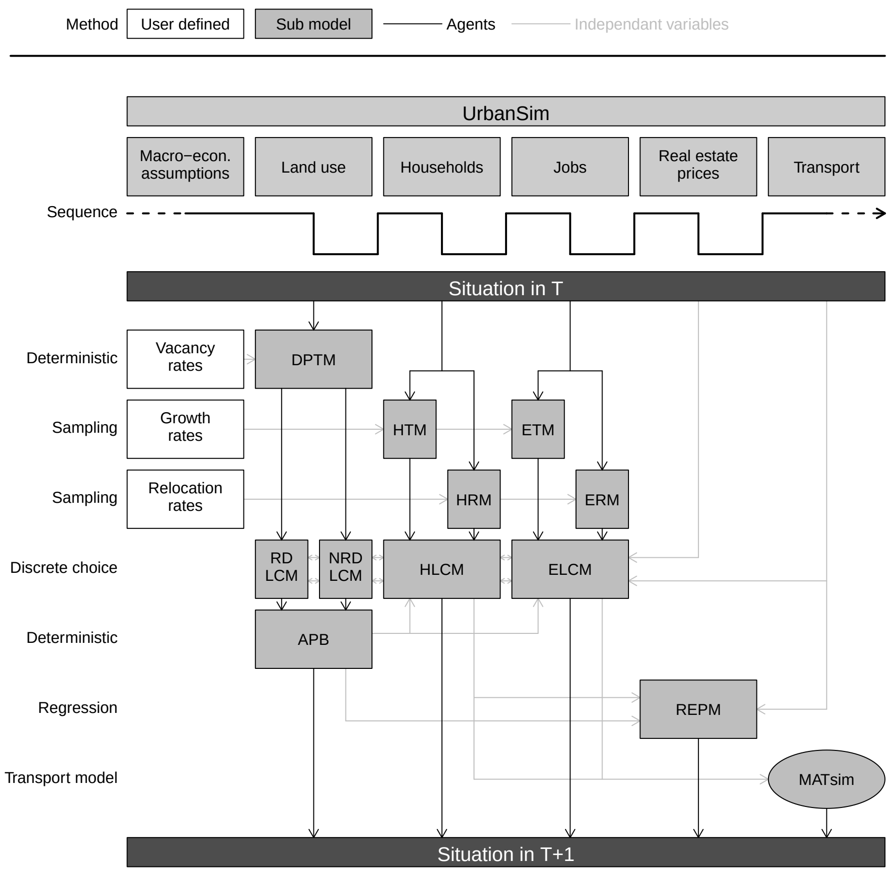

Figure 5.1 – The *UrbanSim* model system

defined by the user, but their value will not be updated by the model system. Note that building types are either residential or non-residential. The former can accommodate households and home-based jobs, while the latter are limited to non-home-based jobs.

Two types of **agents** are represented in *UrbanSim*: households and jobs. Note that firms are not represented in *UrbanSim* and that employment is, therefore, accounted for by these individual jobs. Both categories of agents are attributed to a given building. All **households** belong to the same category but several intrinsic characteristics (see appendix D.1) can be defined. It should be noted, however, that *UrbanSim* does not model demographic or economic evolutions. One household of four persons, having two cars, and earning a total annual income of 50 000 monetary units will, therefore, keep the same characteristics during the simulation period. The only exception is that households can relocate from one building to another. A demographic sub model was developed during the *SustainCity* project (see Turci et al., 2015), but has not been implemented in any operational application of *UrbanSim* (including those of the project).

Jobs are first divided between home-based and non-home-based jobs. The former can include all types of activities, while the latter are further divided among the activity sectors (which are user-defined). Firms are, however, not explicitly modelled. Intrinsic characteristics of jobs are, therefore, limited to the category (i.e. home-based or non-home-based status), the activity sector, and the id of the building where the job is located. This latter variable is, as for households, the only one that can evolve during the simulation period.

Finally, although *UrbanSim* relies on households, *MATsim* operates at the level of the individual. Hence, an intermediary entities can be defined, the persons, allowing the exchange of information between the two models (see section 5.2.1). As indicated by its name, the persons' level consists in a disaggregated representation of every individual composing the households. Personal characteristics (see appendix D.1) can thus be added but are generally limited to the id of the job occupied by the person (if any). As already mentioned, there are no demographic components within *UrbanSim*, meaning that the characteristics of persons are constant over the simulation period, as for households.

5.2.1 Iteration

UrbanSim operates by iteration of one year during which five sequences of sub models are called (Figure 5.1). Before reviewing them, let us note that all sub models within *UrbanSim* correspond to only four different methods:

- **Deterministic:** a sub model that compute a quantity, using a mathematical formula and various inputs;

- **Sampling:** a sub model that select a given number of agents by random sampling;

- **Discrete choice:** a DCM predicting the location choices of agents (see chapter 4 for details on the econometric specification);

- **Regression:** a regression predicting the future level of real estate prices (see chapter 3 for details on the econometric specification).

Buildings

The first sequence ("buildings" on Figure 5.1) aims at predicting real estate developments occurring within the study area. Its principle is summarised by Figure D.1 and consists in the successive cal of three sub models:

1. In the Development Project Transition sub model (DPTM), the vacancy rate (i.e. the number of existing residential units minus the number of households) is compared to the (exogenously defined) average long-term vacancy rate. If the current rate is lower than the target, future development project are created in order to reach the target;

2. The location of these development projects is determined by the residential development projects' location choice sub models (RDLCM), and constrained by development capacities: real estate developments can only occur if the building is not at its (exogenously defined) maximal capacity;

3. Finally, the last sub model of the sequence (Add Project to Buildings) updates the characteristics of the buildings to which a new development project have been affected.

The same procedure is repeated for every building type. Note that on Figure 5.1, NRDLCM refers to the non-residential development projects' location choice sub model. RDLCM and NRDLCM are divided into one specification for each building type.

Households and jobs

The second and third sequences are identical in their principles (Figure D.2 and D.3). Their role is to forecast future location of new or relocating households and jobs, as described hereafter:

1. The Household Transition sub model (HTM) simulates population growth. The total number of households at each iteration is defined exogenously as a macro-economic assumption. Hence, the HTM sub model will first compute the number of new households, equal to the difference between the total defined for t minus the total in $t-1$. Then, it draws the corresponding pool, by random sampling, among existing households. The new households are created by duplicating this sample;

2. The Household Relocation sub model (HRM) selects the moving households, i.e. those that will change their residential location. The number of moving households is computed as the total number in t times the households' relocation rate (an user defined macro-economic parameter);

3. Finally, the Household Location Choice sub model (HLCM) allocates the set of relocating households (new + moving) to available locations.

Note that households' relocation rates can be divided according to age of head and income level. As mentioned above, the procedure is identical for jobs, except that different macro-economic assumptions and specifications of the Employment Location Choice sub model (ELCM) have to be declined for each activity sector. Chapter 4 details the econometric specification of the ELCM. *Stricto Sensu*, the model allocates new or relocating agents to buildings. However, except their type, real estate values and capacities (i.e. the number of households or jobs that they can receive), buildings have no intrinsic characteristics. In many implementations of *UrbanSim* (Waddell et al., 2003, Waddell et al., 2007, de Palma et al., 2015a, and Cabrita et al., 2015), location factors of agents are thus limited to zonal attributes (e.g. income level, population or job densities, car accessibility) for jobs, or to a mix with households' characteristic; see also chapter 6).

Real estate prices

At this point, the iteration is almost completed. The fourth sequence (Figure D.4) only consists in updating real estate prices to the new distribution of agents. Hence, the Real Estate Price sub model (REPM) predicts the price level in $t+1$, using a log-linear regression model (see chapter 3). Note that a different specification of the REPM is used for each building type.

Transport

The fifth sequence (Figure D.5) calls the external travel model (here *MATsim*), to update transport indicators. The reasons for using *MATsim* are given in chapter 1. Although it is a micro-simulation activity-based model able to simulate multi-purpose trips, the coupling plug-in with *UrbanSim* is limited to home to work trips (see Nicolai and Nagel, 2010, Nicolai et al., 2010, Nicolai and Nagel, 2011 and Nicolai and Nagel, 2015). Practically speaking, *MATsim* reads the location of workers and jobs from *UrbanSim*' database to perform the estimation transport fluxes. The following indicators are then exported to *UrbanSim*: (1) at the zone level, the accessibility to jobs by different mode (by foot, by car with and without congestion, and by public transport if any). (2) Home-to-work (and work-to-home) travel times and distances are estimated at the persons level. Note that the accessibility to jobs is computed as a log sum, i.e. a Shimbel index weighted by the number of jobs in each destination.

Additional configuration files are requested by *MATsim*. The transportation network must be stored in external xml file (i.e. not included in the *UrbanSim* database). The road network is accounted for by storing a list of nodes, characterised by their X/Y coordinates, and links that join nodes together. Five characteristics must be given for each link: (1) node of origin, (2) node of destination, (3) maximal travel speeds in m/s, (4) number of traffic lane, and (5) capacity in vehicle per hour. It is possible to define a public transport network, by an origin - destination matrix of travel times and distances, the nodes of the matrix accounting for public transport stops. The implementation of transport scenarios also requires specific configuration files. A cordon fee scenario will, for instance, rely on an additional file to store the list identifying the links of the road network subject to the fee and the characteristics of the fee (i.e. start/end time and amount of monetary units).

5.3 A simple, small-scale, synthetic city for *UrbanSim*

The use of synthetic data sets is frequent in studies on the MAUP, to control the relation between variables (e.g. Fotheringham and Wong, 1991, Amrhein, 1995, and Reynolds and Amrhein, 1998). The purpose of an artificial case study is similar here, since it allows controlling the structure of the city and, therefore, the factors driving agents location choices. Several simplifying assumptions have been made to reduce the number of parameters that control the structure of this synthetic city. The goals of this section are to details these assumptions and to outline the practical implementation of the script generating the database of the synthetic city.

5.3.1 Agents characteristics

We assume that agents are homogenous. All households include three persons (two parents, one children), and both parents work (the complete list of agents' characteristics accounted for by *UrbanSim* is given in appendix D.1). To avoid discrepancies in the transport model, the number of car is set to two. The total annual income (identical for all households) is set to 50 000 euros (close to the average annual income in Belgium according to the 2001 population census). Other characteristics are also based on average values observed in Belgium. All jobs belong to the same employment sector. Given the assumption that all households include two workers, if H is the number of households in t_0, the total number of jobs is equal to $J = 2H$. Note that these jobs are divided among non-home-based jobs (95% of the total) and home-based jobs.

Since agents are homogenous, their creation does not involve a real synthetic population approach (see e.g. Barthelemy and Toint, 2013 for a state-of-the-art application to Belgium, or Farooq et al., 2015b; Farooq et al., 2015a for applications developed during the *SustainCity* project). All fields have identical values (see Table D.1), except the ids (household, building, and zone). Hence, the generation of agents merely involves inputting the correct number of rows (one row per agent) in the households and jobs' tables. The use of homogenous agents is a strong simplifying assumption. It is, nevertheless, consistent with the initial formulation of the Alonso-Muth model, on which the structure of the synthetic city is based (see section 5.3.2).

5.3.2 Spatial structure of the study area

The synthetic city is developed on a featureless landscape. The space is divided into squared grid cells of identical size (the number of rows and columns can be adjusted by the user). Three parameters must be given to set up that grid: the number of rows and columns, and the surface of a cell. The only characteristic of a grid cell is, therefore, its Euclidean distance to each CBD (whose location and relative size must be exogenously provided). However, the mathematical form of the synthetic city, i.e. the functions allowing to compute the initial number of agents per zone, is mostly the result of trials and errors. The reasons are detailed hereafter.

Methodological constraints

When building the synthetic city, we decided that the utility function of households should follow the principles of the Alonso-Muth model, i.e. to (a) increase with the accessibility to jobs and (b) to decrease with the real estate prices. These real estate prices should, in turn, be proportional to both the households and population density. However, contrary to a pure theoretical city approach (e.g. Schindler and Caruso, 2014, Delloye et al., 2015; see also section 7.4.1), this utility function is not defined explicitly in *UrbanSim*, but results from the estimation of the households' location choice model. Practically speaking, the utility level of households in each zone is determined by the variables included in this sub model, and by the sign of their parameter estimates.

Which means that we had to define the initial spatial structure of households and jobs without knowing if the estimation of the location choice and real estate price sub models will produce parameter estimates of the desired sign. Therefore, rather than relying on theoretical density functions such as Clark (1951)' law (a negative exponential; see also Newling, 1969; Batty and Longley, 1994), we had to proceed by trial and error. The mathematical form finally selected (see equations 5.1 and 5.2) allows having the desired direction of the feedback effects on the utility level[1]. It consists in an inverted logistic curve for the population density, and to a negative exponential for jobs. Although somewhat different from above-mentioned theoretical densities' functions, this choice implies that jobs are highly concentrated close to the CBD, while house-

[1]Testing different mathematical form of the synthetic city is a burdensome task. It involves generating a complete database, then to import into *UrbanSim* (which cannot be fully automated), and to estimate manually the econometric sub models. Hence, even if different values of the distance - decay parameter (β) had been tested, no full sensitivity analysis was performed.

holds sprawl more towards the suburbs. Both characteristics are consistent with the Alonso-Muth model.

Mathematical form of agents' densities

The mathematical functions controlling density take thus the form of an inverted logistic curve for households (equation 5.1), and to an exponential decrease is for non-home-based jobs (equation 5.2). They allow determining the attraction potential of a zone i, noted P_i. In equations (5.1) and (5.2), α denotes the relative size of each CBD (see section 5.5), β the distance-decay parameter (set to one here), and n the number of CBDs.

$$P_i(h) = (\sum_{i=1:n} \alpha_n \frac{1}{1 + e^{-\beta_h d_{in}}})/n \tag{5.1}$$

$$P_i(j) = (\sum_{i=1:n} \alpha_n e^{-\beta_j d_{in}})/n \tag{5.2}$$

Potential $P_i(h)$ and $P_i(j)$ are then re-scaled between 0 and 1. The number of households h_i and non-home-based jobs j_i in a zone i in t_0 is thus equal to $HP_i(h)$ and $0.95JP_i(j)$ (rounded to the closest entire value). Home-based jobs are distributed between zones proportionally to the household' density.

Land use and buildings

In a zone version of the *UrbanSim* model, agents are located in buildings, which in turn are situated in one given zone. The building types are limited to *Houses* and *Offices*. They are assumed to be mono-functional, i.e. purely residential or non-residential. Their characteristics (number of residential units, of non-residential square feet, and average value) depend on the number of agents per zone.

The following assumptions are made for residential buildings (*Houses*) in t_0. First, the number of residential units (i.e. existing dwellings) for an house located in the zone i is equal to $ru_0 = h_i(1+vr)$ with vr the average, long-term, vacancy rate (set here to 10%). The residential units capacity ru_c is equal, for all *Houses*, to the maximal value of ru_0. Practically speaking it means that the residential developments' capacity is null in the CBD and increase with the Euclidean distance to the CBD. The identical value of the residential units capacity for all zones reflects the absence of any land-use planning constraints. Note that since the utility level of households is higher close to the CBD, we have to define a threshold value preventing all new or relocating households to locate in the CBD. An additional capacity of zero was the simplest definition for that limit.

Finally, the average value per residential unit (p_i^{ru}) is defined by equation 5.3, where $\mu(h_i)$ denotes the average number of household and C_r a constant. Note that C_r is set to 50 000 here (in order to have real estate price close to values observed in Belgium) but has no practical importance. Since incomes are uniform among households, only the spatial variations of the residential prices matter in their location choices.

$$p_i^{ru} = \frac{h_i}{\mu(h_i)} \times C_r \tag{5.3}$$

Non-residential buildings (*Offices*) are similarly characterised by the existing floor space for jobs, the floor space capacity (existing + developable) and their average price. Note that this latter characteristic is here the value of one square meter of floor space, not (as for *Houses*) the price of the entire building. This is an hard-coded assumption in *UrbanSim*' source code, see Waddell (2000). The non-residential surface in t_0, noted nr_0, is equal to $20j_i(1 + vr)$. In other words we assume that each job require a surface of 20 square meters. The non-residential surface capacity (nr_c) is equal to the maximal value of nr_0. The real estate prices in t_0 are given by equation 5.4, where $\mu(j_i)$ denotes the average number of non-home-based jobs per zone, and C_{nr} is a constant (set to 100 here).

$$p_i^{nr} = (\frac{h_i}{\mu(h)} + \frac{j_i}{\mu(j)}) \times C_{nr} \tag{5.4}$$

Finally, note that the buildings are assumed to be mono-functional, meaning that the non-residential surface (both existing and potential) is set to zero for all *Houses*. Non-residential buildings, conversely, does not include any residential unit.

5.3.3 Macro-economic assumptions

The macro-economic assumptions represent the steady state of the studied urban area. They are, therefore, user-defined and specific to each case study. The only assumption here is that the population growth is linear, meaning that H_n, the number of households in t_n is equal to $H_0(1+g)^{n-1}$ with n the number of year and g the population growth rate. The control totals for jobs are derived from these values for households, following the rule that $J = 2H$. Another simplifying feature is that the (user-defined) relocation rates are identical for households and jobs and constant since agents are homogenous.

5.3.4 Transportation network

A road network connects the centroid of each zone to the centroid of all adjacent zones, on a von Neumann neighbourhood and in both directions (i.e. A to B and B to A). The length of each link is the Euclidean distance between the two centroid. Maximal speed is set to 13.88 m/s (50 km/h, the maximal authorised speed in urban areas in most European countries), and the capacity of each lane (one in every direction) to 500 vehicles per hour.

We decided not to include any public transport system, since it would have required numerous assumptions about the location of the public transport stops, and on the travel times. Moreover, the accessibility to jobs by public transport would only influence the location choices of households if this variable is included in their utility function (see Table 5.2). We recognise that this methodological choice constitutes a strong limitation. Nevertheless, given the limited additional indicators that a public transport network would have offered (limited to travel times, since technical problems were encountered with the model choice component; see chapter 6), it seemed not to be worth the trade off.

5.3.5 Practical implementation

The synthetic city is generated by a script, written in R. The inputs are limited to a csv file storing the user-defined parameters (see section 5.3.2). The outputs consist in a database (csv files) consistent with the requirements of zone version of *UrbanSim* (see appendix D.1). Table D.3 summarises the sequence followed by the script. Note that different actions are required after the generation of the database before being able to run simulations. Overall, these steps are the following:

1. Input user-defined parameters within a csv file;
2. Run the "synthetic city" script within R (see Table D.4);
3. Upload the different tables into a SQL database (this step is required to ensure that *UrbanSim* store the different fields in the good encoding, i.e. integer, float, or string - an additional R script allows automating this procedure);
4. Upload the SQL database to OPUS using the built-in functionnalities;
5. Define and estimate the econometric sub models (REPM, HLCM, ELCM, RDPLCM, NRDPLCM - this step has to be done manually);
6. Run the *MATsim* to compute initial values of travel and accessibility indicators (practically speaking, it consists in running *UrbanSim* for one iteration with all other sub models deactivated);

7. Run complete simulations within *UrbanSim*

5.4 Mono centric configuration

This first implementation of the synthetic city is used to assess both the influence of the *scale* and *boundary effects*. This section consists, however, in an exploratory analysis. There is two reasons for that. First, the development of this synthetic case study is mostly the result of trial and error, leading to a somewhat non careful definition of the experiment plan. Secondly because several components of the model system are neutralised, to keep the feedback effects as limited as possible. The assessment of spatial bias is, therefore, essentially based on pairwise comparisons.

Note that we assume a mono centric synthetic city, and an higher utility level close to the CBD for both households and jobs. The reason of this choice is to reduce the influence of variations of the *spatial extent*. Since the aim of this work is to assess the sensitivity of *UrbanSim* to spatial bias, we feel necessary to rely on a case study where the influence of these biases should be limited.

5.4.1 Methodology

Structure of the synthetic city

The synthetic city is assumed to be a grid of 45×45 zones (the central one being the only CBD). The surface of each of the 2 025 zones is of 2 km^2. The initial number of households is set to 405 000 in order to have an average population density of 100 households (or 300 inhabitants) per square kilometre. Given the parameter used in the potential equations (5.1 and 5.2), it gives a maximal number of households per zone of 350, or 385 residential units (with the 10% vacancy rate). Macro-economic assumptions are the following: we assume a constant growth rate (of 1% per annum) of the number of households. The relocation rate is set to 10% per annum. Finally, the target vacancy rate for residential buildings has been arbitrarily fixed to 10%.

In order to simplify the behaviour of the model, the evolution of the jobs is not accounted for. Growth and relocation rates are set to zero for this category of agents. The corresponding sub models are thus neutralised during one iteration of *UrbanSim*. Note that the existing stock of dwellings is sufficient to accommodate the total number of households at the end of the simulation period, allowing to neutralise also the sub models simulating the development of new residential or non-residential capacities.

Case studies

Three case studies are defined (Figure 5.2). To assess the influence of the *scale effect*, two nested BSU levels are compared: (a) *Reference* and (b) *Aggregated*, which consist in a simplification of the city into a 15 by 15 grid, meaning that each large BSU is constituted of a square of nine initial zones (for a total of 149 large BSUs). Re-aggregation procedure between the *Reference* and *Aggregated* case study is limited to land-use (i.e. buildings and zones): characteristics of aggregated zones or buildings are computed using initial zones or buildings values, by sum (surface, number of existing and potential residential units) or mean (average value per unit, distance to the CBD). For agents, the re-aggregation procedure consists in the replacement of their initial id(s) by the id of the aggregated zone or building to which the agent belongs. Note that to avoid shape problems (see section 5.4.2), it has been decided to use a subset of the initial synthetic city, consisting of a circular study area. The extension is based on the *Aggregated* case study and includes all large BSU having their centroid in the circle inscribed in the initial square.

The *boundary effect* considers the *Reference* case study as the complete extension of the study. A *Subset* case study is generated by selecting all the zones for which the Euclidean distance between their centroid and the centroid of the CBD is less or equal to an arbitrary selected threshold of 60% of the maximal distance (see Figure 5.2). The sub setting procedure consists in extracting from the original database the agents and buildings that belong to one of the zone of the subset. Since the BSU level does not change, no re-computations of buildings or zones characteristics are necessary. Control totals are adapted: the share of households in the *Subset* relative to the total for the *Reference* case study is computed, and the same ratio is used for the population growth in the subset (i.e. final - initial number of households). Descriptive statistics for all case studies are presented in Table 5.1. Figure 5.3 shows the main characteristics of the synthetic city in t_0.

Scenarios

Three scenarios are implemented on each case study (ending up with nine pairs case study/scenario). First, the *Baseline* scenario assumes a linear evolution of the synthetic city, without external shock. Secondly, a transport infrastructure improvements scenario (*Highways*) has been defined as follows: the replacement of all roads between the CBD and the BSUs located on the north, east, south and west by express roads. The attributes of the corresponding links in the network used by *MATsim* have an average speed of 25 m/s (90 km/h), two circulation lanes and a capacity of 3 200 vehicles per hour. Figure 5.4a shows the relative variation of the car accessibility to jobs induced by the implementation of this scenario. Thirdly, Urban Growth Boundaries (*UGB*) are defined by increasing the provision of residential units in the BSUs close to the CBD (i.e. at less than 40% of the maximal distance to the CBD) and decreasing it in the BSUs located further away (Figure 5.4b). Note that the total number of residential units remains identical between scenarios. Both *Highways* and *UGB*

			Case study		
			Reference	Aggregated	Subset
BSU			1341	149	861
Agents	Households		321 160	321 160	236 284
	Jobs		610 191	610 191	433 828
Variables	Hab/km^2	Min.	75.0	78.6	101.5
		Mean	119.7	119.7	137.2
		Max.	176.0	175.3	176.0
	Jobs/km^2	Min.	169.0	172.9	198.0
		Mean	227.5	227.5	251.9
		Max.	426.0	397.6	426.0
	Car accessibility to Jobs (Logsum)	Min.	7.70	8.84	8.15
		Mean	9.18	9.63	9.30
		Max.	9.68	10.20	9.74
	(Log of) Houses price	Min.	9.19	9.50	10.71
		Mean	11.41	11.43	12.14
		Max.	14.93	14.73	14.93

Table 5.1 – Mono centric case studies

scenarios take place in t_0, meaning that the changes in the baseline situation occur before the start of the simulations.

Note that these scenarios have been designed to trigger a response of the model system rather than to mimic land-use policies currently implemented in metropolitan area. In particular, the implementation of the *Highways* scenario should have a direct effect on households' location choices, since it affects one of the independent factors driving their utility level (see Table 5.2). An improvement of the existing road network is a simplifying assumption but is also, on a practical point of view, the more straightforward way to implement a change in the car accessibility to work. Obviously, for real-world case studies, the transport scenarios implemented nowadays would rather consist in restriction on the road network (e.g. cordon toll) or the development of alternative mode (see chapter 6). The effects of the *UGB* scenario are, on the contrary, indirect. The utility level of households is not affected by its implementation, but the available location yes, which may influence the final situation predicted by *UrbanSim*.

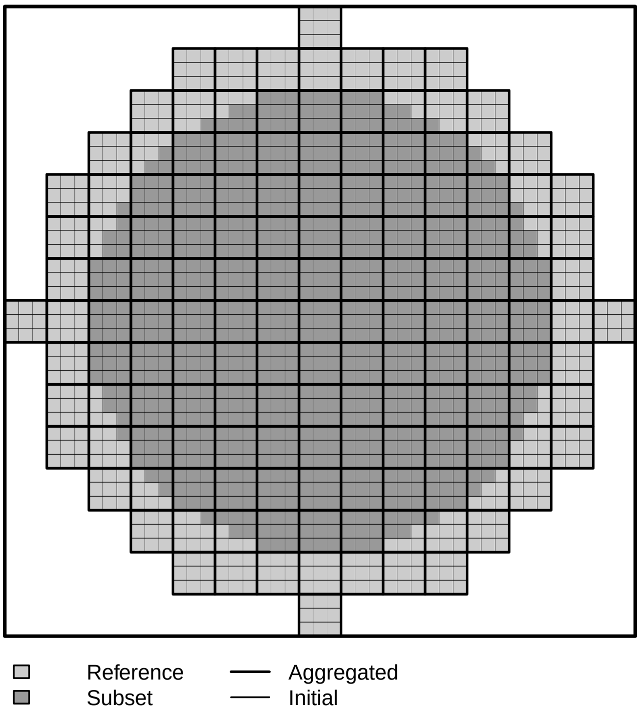

Figure 5.2 – Mono centric case studies

Calibration of econometric sub models

The aim is to keep the number of feedbacks within the synthetic city as low as possible. Econometric sub models have thus been limited to variables consistent with the Alonso-Muth model. Location choices of households are assumed to depend only to the (log of) job density and houses prices, and on the car accessibility to jobs. Parameter estimates for both case studies are given in Table 5.2. This set of independent factor is implemented for two reasons: (a) their relevance for households' location choices (e.g. Anas, 1982, Fujita, 1989), and (b) because the values of these variables are updated at each iteration of the model, which allows taking into account the evolution of the synthetic city in location choices. Location choice sub models are estimated within *UrbanSim*, using a stratified sample of 10% of the agents located in each zone. A steady-state is, therefore, assumed for the synthetic city.

Population and jobs densities are thus the only factors affecting houses price. Note that the latter is constant over the simulation period, since the evolution

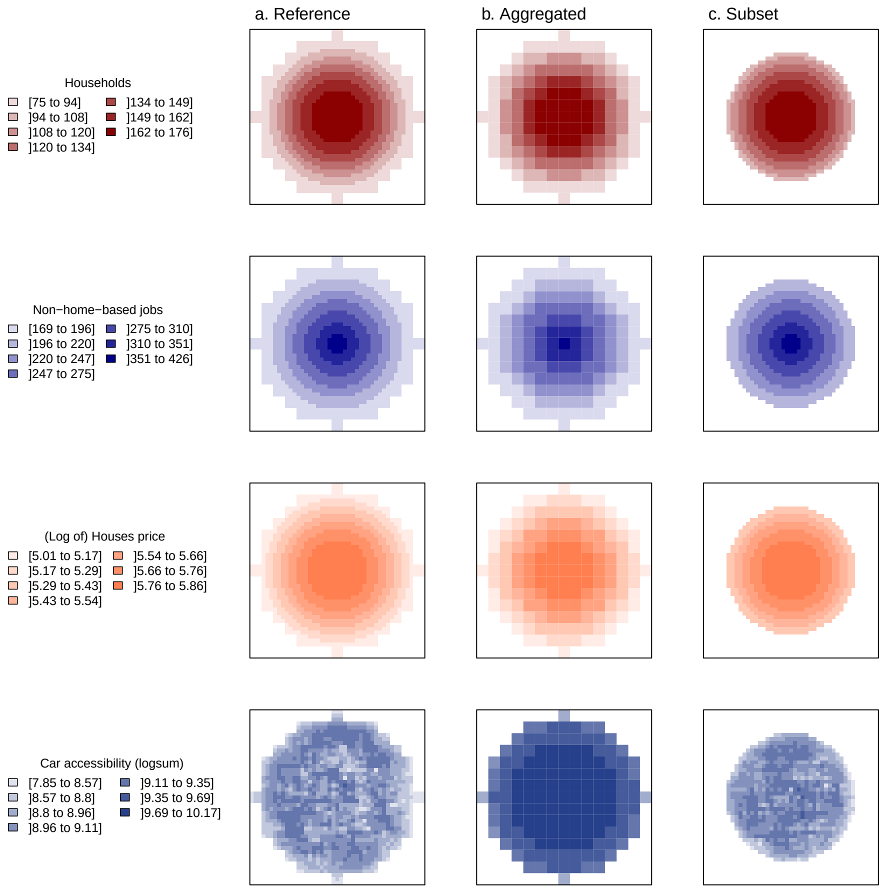

Figure 5.3 – Characteristics of the mono centric case studies (values in t_0)

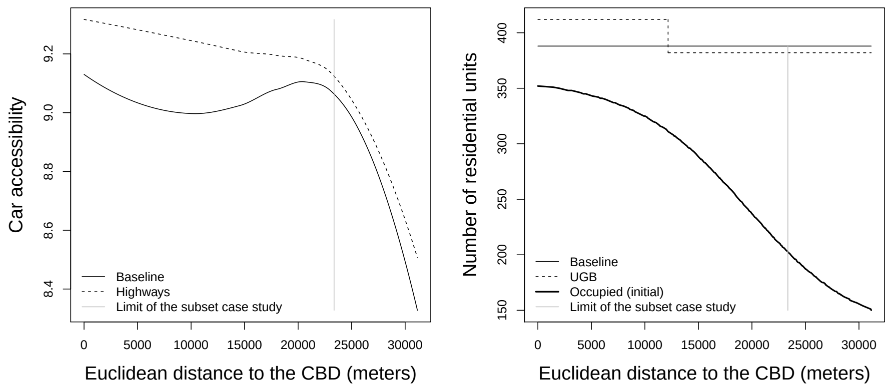

Figure 5.4 – **Scenarios** (variations induced by the implementation of the *Highways* and *UGB* scenario; *Reference* case study)

of jobs is not modelled. However, limiting the real estate price sub model to the population density makes the system perfectly singular, preventing the estimation of this sub model, reason why the job density is included. Even if not satisfactory from an intellectual point of view, this choice has no practical influence on the outputs. Since the BSUs are featureless, there was no need to include local amenities' indicators. Parameter estimates are given in Table 5.3. Note that the algorithm generating the synthetic city precisely assumes that houses' prices are a function of population and job densities (see section 5.3), explaining the nearly perfect adjusted R^2 values.

The feedback effects occurring in the model are thus limited: (1) new households locating in one zone will increase houses' prices in that zone, causing a decrease of its utility level. (2) Relocation of households may as well affect the car accessibility to jobs, although on a less predictable way. Note that the *Highway* scenario will also affect this variable.

Simulations

The simulation period lasts 20 years (or the same number of iterations of *UrbanSim*). Thirty runs are performed for each combination case study/scenario to cope with the stochastic variations of the model (random sampling of relocating agents), following the recommendations of Wegener (2011a). Configuration of the *MATsim*/*UrbanSim* interface uses the default parameters presented in Nicolai and Nagel (2010). Iterations of *MATsim* being more computationally intensive than for *UrbanSim* (on the configuration used for simulations, about two hours compared to a few minutes), the travel model is estimated every five

Variables	Case study Reference	 Aggregated	 Subset
(Log of) Job density	5.39 (0.13)	132.08 (2.13)	4.73 (0.15)
(Log of) House price	-0.59 (0.11)	-85.99 (1.56)	0.28 (0.16)
Car accessibility	-0.11 (0.03)	6.78 (0.2)	-0.1 (0.05)
AIC	192 856	123 637	145 649
Likelihood ratio	0.10	0.42	0.07
n	31 503	31 503	23 219

Table 5.2 – Households location choice sub model (between brackets: standard deviation; all parameters significant at $\alpha \leq 0.05$)

Variables	Case study Reference	 Aggregated	 Subset
Constant	4.509 (0.002)	4.51 (0.007)	4.59 (0.001)
Pop. density	0.01 ($6\text{x}10^{-5}$)	0.01 ($1\text{x}10^{-4}$)	0.008 ($3\text{x}10^{-5}$)
Job density	-$6\text{x}10^{-4}$ ($2\text{x}10^{-5}$)	-$7\text{x}10^{-5}$ ($6\text{x}10^{-6}$)	-$3\text{x}10^{-4}$ ($1\text{x}10^{-5}$)
R^2 (adjusted)	0.99	0.99	0.99
n	1341	149	861

Table 5.3 – Real estate price sub model (for houses; between brackets: standard deviation; all parameters significant at $\alpha \leq 0.05$)

iterations of *UrbanSim*. Overall, the sequence of sub models executed during an iteration is the following (see also appendix D.2)[2]:

1. Households transition sub model (HTM);
2. Households relocation sub model (HRM);
3. Households location choice sub model (HLCM);
4. Real estate price sub model (REPM);
5. For iterations 1, 5, 10 and 15: External travel model (TM, here *MATsim*).

No calibration procedure is performed *per se*. This step usually consists, for LUTI models, in running the model in-between the base year and a time

[2]The following sub models of *UrbanSim* are neutralised: Development project transition sub model, Residential and Non-Residential development location choice sub models, Add project to buildings sub model, Employment transition sub model, Employment relocation sub model, Employment location choice sub model, Distribute unplaced jobs sub model and Refinement sub model.

step for which observed data are available, then compare it with the forecasted situation at this time step (Wegener and Furst, 1999a). The synthetic city being generated for only one time-step, it is impossible to perform such analysis here. Moreover, the aim of this section is to assess the variations in the response of the model between different size of BSU and study area, not to make actual predictions.

Finally, the variations observed between case studies and/or scenarios are significant only if they exceed the inter-run variations (Table 5.4). Note that they are limited in magnitude for *Reference* and *Subset* case studies, and extremely similar between scenarios. Larger variations are found for the *Aggregated* case study.

	Scenario		
Case study	Baseline	Highway	UGB
Reference	3.23	3.20	3.31
Aggregated	24.99	24.93	24.48
Subset	2.31	2.33	2.40

Table 5.4 – **Inter-runs variations** (average over the 30 runs; standard deviation of the difference from mean in the final share of households per zone)

5.4.2 Results

Assessment of the influence of the *scale* and *boundary effects* focus on the differences observed in the final distribution of households. Without further notices, all indicators are given for t_{20} (last year of the simulation period) and that the evolution is computed as the value in t_{20} minus the value in t_0. It should be noted that the results presented hereafter have limited meaning from economic or geographic point of view. The validity of the variations observed (e.g. a larger urban sprawl for the *Aggregated* case study) is, therefore, limited to our synthetic city. Nevertheless, they will hopefully allows a better understanding of the mechanisms driving the sensitivity of the outputs of *UrbanSim* to the *scale* and *boundary effect*. This is the main purpose of our experiments. Figure 5.5 shows the evolution of the number of households per zone between t_0 and t_{20}.

Scale effect

Comparing the *Reference* and *Aggregated* case studies allows estimating the influence of the *scale effect*. Large differences are observed in the final number of households per zone, varying (as *Aggregated* minus *Reference*) from -1 873 to +1 618. Relative variations go from -67 to +86%. Hence, the spatial structure of the population changes dramatically in t_{20} for the *Aggregated* case study (Figure 5.6; left). Households relocate from zones at a medium distance from the CBD to zones situated either (1) in the periphery of the study area or (2) close to the CBD.

Boundary effect

The *boundary effect* is assessed by comparing the *Reference* and *Subset* case studies. Over the 30 runs, the *Reference* case study predicts a number of households within the subset in t_{20} that vary from 278 724 to 279 355. This is significantly less than the final number of households for the *Subset* case study (285 458). (Note that this latter final population was defined *a priori* by the macro-economic assumptions). Relative variations of the number of households are, therefore, mostly positive (Figure 5.6; right). They vary from -13 to +19 (+7 on average). Relative variations go from -6.7 to +3.8%. A strong spatial structure appears, with larger differences in "dense suburbs" while low or even negative differences are observed in the periphery of the *Subset* case study. The presence of negative differences, despite the lower final number of households for the *Reference* case study, suggests that variations would have been larger if the totals were equal.

Influence of the scenarios

Variations due to the *scale* or *boundary effects* will only constitute an issue if they are (1) larger than inter-runs variations and (2) comparable in magnitude to the influence of potential scenarios (*Highways* and *UGB*). Let us start with the *Highways* scenario. Relative variations with the *Baseline* in the final number of households per zone are comprised between -4.4 and +4.7% for the *Reference* case study. They are of -19 to +35% for the *Aggregated* and of -3.5 to +2.1% for the *Subset* case studies. No spatial structure emerges, whatever the case study (Figure 5.7; upper row), and such values do not exceed inter-runs variations. The improvements of the transportation network implemented by the *Highway* scenario have thus no significant impact on the evolution of the city. The main reason being that the changes in the car accessibility to jobs are insufficient to impact the location choices of households.

Figure 5.7 (lower row) also shows the variations induced by the implementation of the *UGB* scenario. For both the *Reference* and *Subset* case study, the final distribution of households is similar to those of the baseline. An upward shift is observed for the *UGB* in central areas (on average, +6.1% for the *Reference* case study and +5.8% for the *Subset*). A corresponding downward shift is observed in peripheral areas (of, respectively, -1.8% and -2.7%). Similar observations are made in central areas for the *Aggregated* case study, but the situation is more confused in peripheral areas. Note that the increase observed

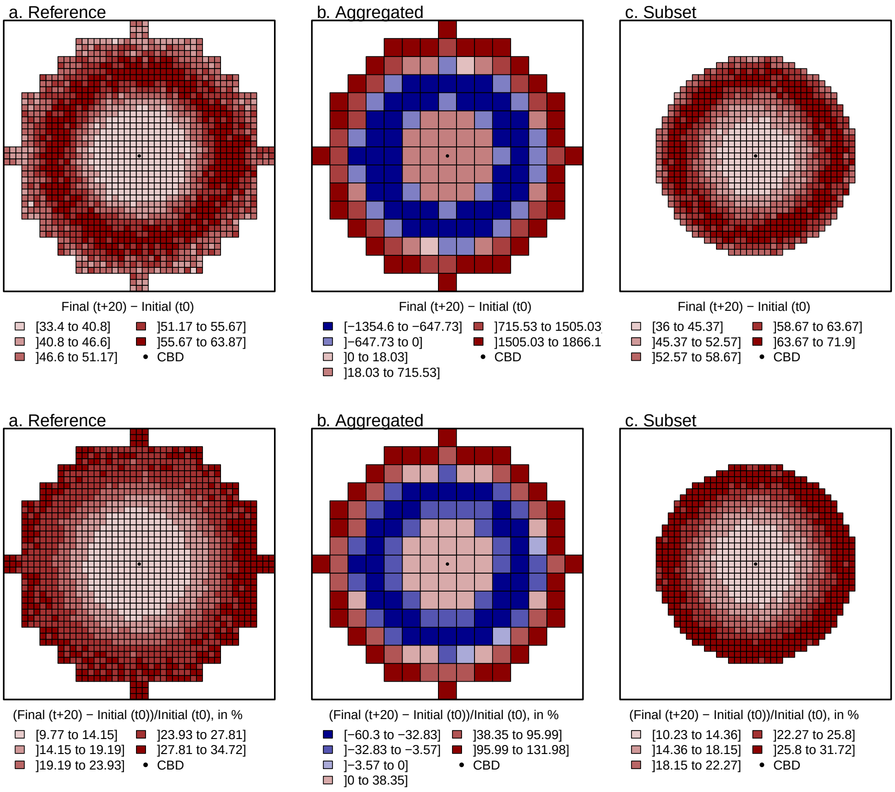

Figure 5.5 – Evolution of the number of households per zone (upper row: absolute variations; lower row: relative variations; discretisation: Jenks)

in central areas has a magnitude equal to the changes implemented by the *UGB* scenario (i.e. 24 additional residential units per zone).

5.4.3 Discussion

Given the simplifying assumptions made on the structure of the synthetic city, and the neutralisation of several components of the model system, it is not possible to derive general implications, on a geographic point of view. Nevertheless, the present case studies are useful to assess the mechanisms driving the sensitivity of *UrbanSim* to spatial biases, which can be summarised as follows:

1. Parameter estimates of econometric sub models vary with the *scale* and *boundary effects*;

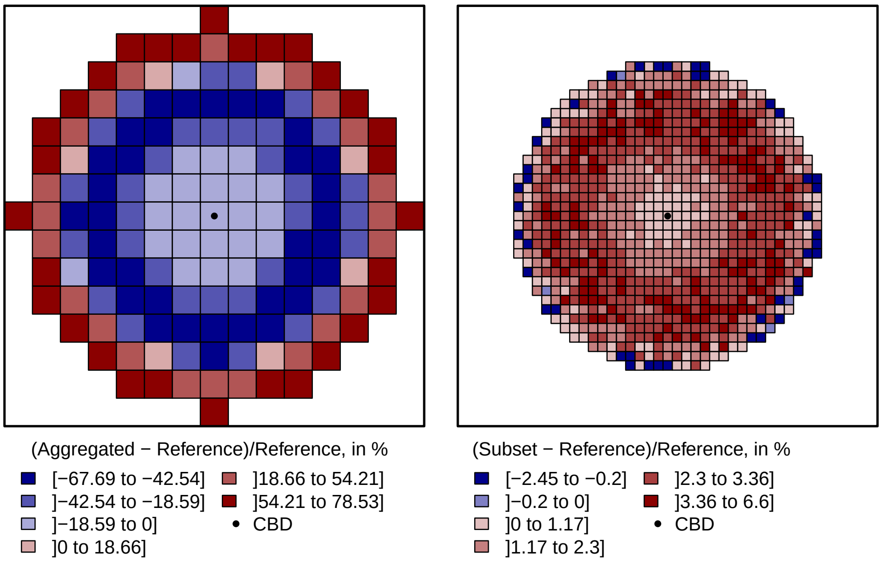

Figure 5.6 – Sensitivity to the spatial extent and resolution (variations of the final number of households per zone; left = *scale effect*; right = *boundary effect*; discretisation: Jenks)

2. These variations lead to differences among case studies in the utility level perceived by households (see chapter 4);
3. The probability of location in each zone is, therefore, affected;
4. Hence, ultimately, *UrbanSim* will locate new or relocating households in different places.

Step one is demonstrated in chapter 3 for the real estate price sub model and 4 for the location choice sub models. For the synthetic case study, see Tables 5.2 and 5.3. Chapter 4 also assesses steps two to four. The present chapter provides two additions. First, the experiments are conducted with *UrbanSim* itself, rather than by replicating its internal principles in an external framework. Secondly, multiple iteration are performed. Feedback effects are thus present, which may either mitigate or reinforce the influence of the variations of parameter estimates.

Variations of the probability of location

Let us first examine these four steps without considering feedback effects. The variations of parameter estimates of the households' location choice sub model

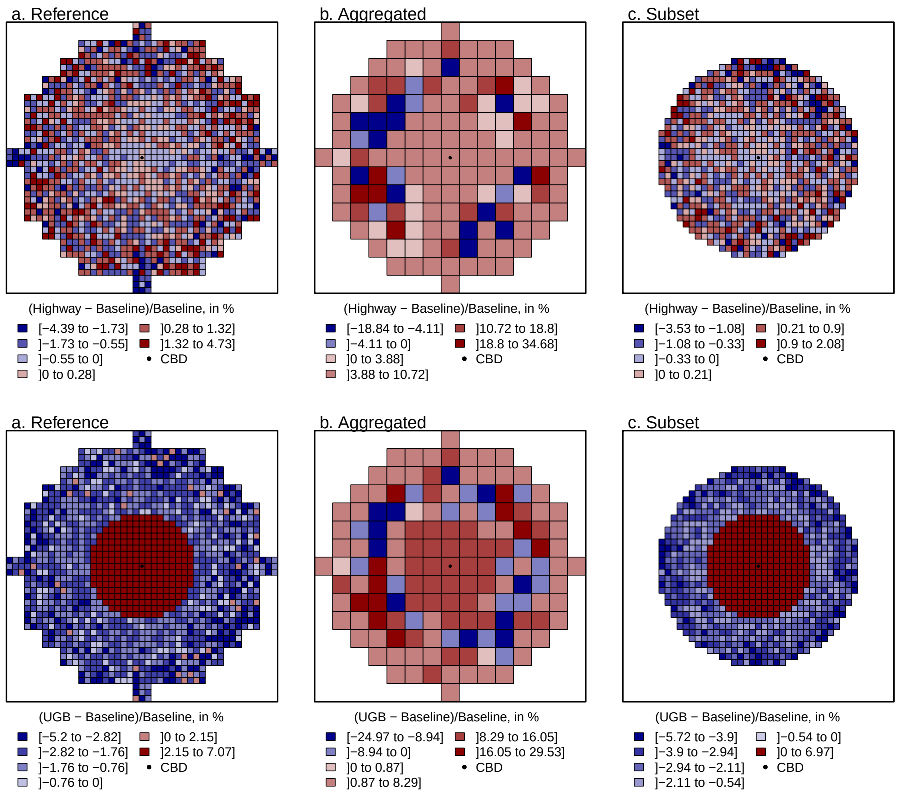

Figure 5.7 – Influence of the scenarios (upper row: *Highways*; lower row: *UGB*; discretisation: Jenks)

(Table 5.2) are consistent with those observed in chapter 4. Change of signs can be observed between case studies. Note that parameter estimates are more stable for the real estate price sub model (Table 5.3). Hence, the utility level of each zone is also different from one case study to another. The potential influence on the outputs of *UrbanSim* is, however, more straightforward to assess by using the variations of the predicted probability of location (as demonstrated by chapter 4).

Figure 5.8 shows these predicted probability of location in t_0. Since the sum of the probability of location is equal to one, the area under each curve is identical for all case studies. Therefore, Figure 5.8 shows that a larger concentration of households close to the CBD can be expected for the *Aggregated* case study (compared to the *Reference*), and a lower one for the *Subset*. Moreover,

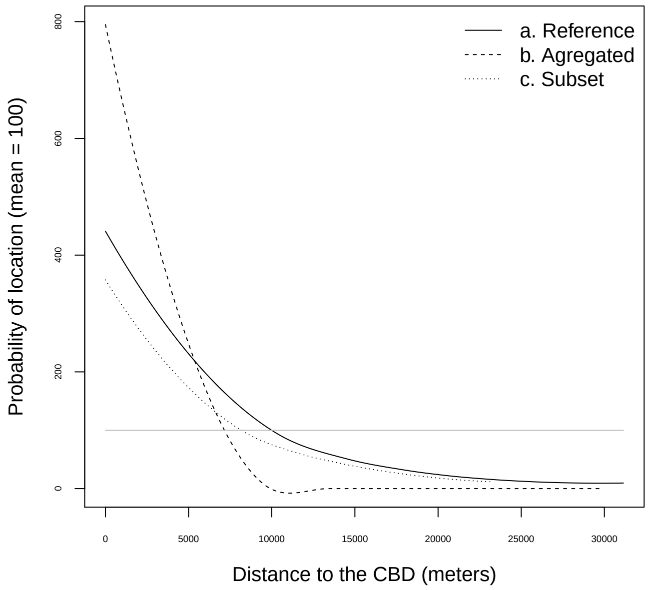

Figure 5.8 – Probability of location of households in t_0 (curves: moving average)

for the *Aggregated* case study, the probability of location is slightly smaller for zones between 10 and 12 km from the CBD than for zones further away.

The observed differences in the number of households per zones for the *Reference* and *Subset* case studies (Figure 5.6; right) are consistent with these variations of the probability of location. The link is less clear in the *Reference* versus *Aggregated* situation (Figure 5.6; left). A larger concentration of households does indeed take place close to the CBD for the *Aggregated* case study, but the largest decreases (compared to the *Reference*) are observed for zones ranging from 15 to 20 km from the CBD (Figure 5.6). This distance is sensibly higher than the 10 to 12 km bandwidth that exhibits maximal differences in the probability of location (Figure 5.8).

Hence, the experiments conducted here validate the theoretical mechanisms by which the outputs of *UrbanSim* are affected by changes in the *spatial extent* or *resolution*. They also prove that the initial (i.e. in t_0) probability of location is useful to predict variations due to the *scale* and *boundary effects*.

Feedback effects on location choices

This initial probability of location is, however, not totally sufficient. The reason is that the variations of parameter estimates imply that the feedback effects within the model system will change in intensity or even direction among case studies. Two of these feedback effects exists here. First, an increase in the population or job density in one zone may generate congestion, reducing the car accessibility to jobs and, therefore, affecting the utility level of the zone on the next iteration. However, the car accessibility to jobs varies here of less than 1% between t_0 and t_{20} (without any spatial structure and whatever the case study).

The main feedback effect is, therefore, the evolution of real estate prices (Table 5.5). Its spatial structure is similar between case studies, with a larger increase close to the CBD. This was expected, given that houses prices are proportional to population density (Table 5.3). Given the sign of parameter estimates (see Table 5.2), an increase of these prices results, however, in a decrease of the utility level for the *Reference* and *Aggregated* case studies, but to an increase for the *Subset*.

Figure 5.9 show the initial probability of location in t_0 and the variations between t_0 and t_{20}. As expected, it decreases close to the CBD for the *Reference* case study, but increases for the *Subset*. Which concurs reducing the differences in the final share of agents close to the CBD that would be observed if the probabilities of location were constant throughout the simulation period.

Note that the *Aggregated* case study is not represented on Figure 5.9. The probability of location in the central zone is above 99%, meaning that absolute differences are maximal there, but that huge relative variations are observed in peripheral areas. Besides the cartographic issue (discretisation), the spatial structure is not consistent with the greatest variations of the share of households per zone (Figure 5.6).

The outputs of the model are consistent with both (a) the initial structure of the synthetic city and (b) the evolution of the utility level perceived by households. The feedback effects contribute here to mitigate the variations that should be observed in the outputs of *UrbanSim*. There is, however, no indication that this result is valid for other case studies. To sum up, since the final probability of location cannot be computed prior to running the simulations, the variations in the outputs of LUTI models for different *spatial extent* and *resolution* are likely to remain unpredictable.

Consistency and limitations

Wegener (2011a) shows that the magnitude of inter-runs variations is a function of the ratio between choices (i.e. the number of agents who relocate during an iteration) and alternatives (the number of zones where they can relocate). This ratio is of 28 for the *Reference* case study and 33 for the *Subset* (with a slight

Case study	t_{20} minus t_0, in % Min.	Mean.	Max.	ρ
Reference	16.8	28.5	41.3	-0.89***
Aggregated	18.0	29.8	46.3	-0.99***
Subset	18.1	27.5	38.5	-0.96***

Table 5.5 – Variations of houses' prices (*Baseline* scenario; ρ = Pearson Product-Moment Correlation with the euclidean distance to the CBD)

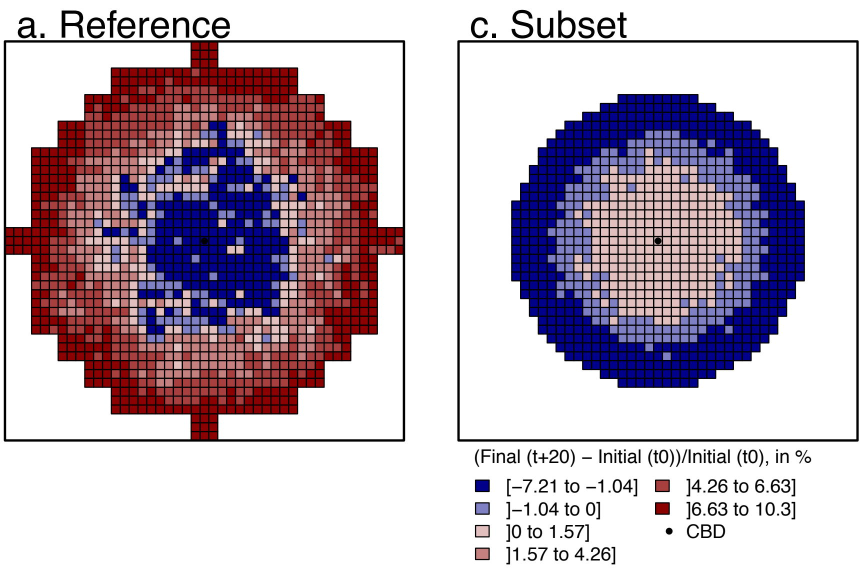

Figure 5.9 – Variations of the probability of location of households in t_0 (discretisation: quantiles).

increase over the simulation period due to the population growth). For such values, the inter-runs variations observed here (see Table 5.4) are lower here than those found by Wegener, 2011a. It may be due to the strong spatial structure of the utility level (Wegener, 2011a uses uniform utilities). Hence, although the variations between case studies and scenarios are of the same order of magnitude than these inter-runs variations, it can be assumed that they represent actual variations rather than noises.

The outputs of the *Aggregated* case study raises questions. The relocation process visible in Figure 5.6 is consistent with the evolution of the utility level. Its intensity is, however, greater than any of the variations observed in other case studies. Inter-run variations are also far larger for this case study (Table 5.4), although the larger choices on alternatives' ratio should have reduced it (as in Wegener, 2011a). A potential reason is that the relocation rate is identical in-between *Reference* and *Aggregated* case studies. The probability that one household decides to move from one zone to another is thus equal, although the size of the BSU varies, resulting in an ecological fallacy problem (Robinson, 1950). It is, however, unclear if this issue is enough to generate such a large relocation phenomena.

Operational implications

Variations induced by the *boundary* effect are limited in magnitude for our case study, since the zones not included in the *Subset* have a low utility level, but remain significant. This limited influence is, however, due to the strong mono-centric structure of the synthetic city. Most metropolitan areas nowadays show as sub urbanisation process, linked to the preferences of households (see Anas, 1982, Fujita, 1989). Moreover, heterogeneity among households has not been considered. Again, location preferences are commonly different between socio-economic groups, and spatial segregation occurs (see e.g. Anas, 1982, Fujita, 1989, and Zenou, 2009). The influence of the choice of the study area on LUTI models' outputs is, therefore, likely to be larger for real world applications. This question of the delineation of the study area and of the relation with the "rest of the world" will be analysed in a more detailed way by section 5.5.

Even for a very simplified case study as the one used here, it is not straightforward to predict the variations in the outputs of *UrbanSim* for different *spatial extent* or *resolution*. Comparing the initial probability of location, as we did here, can help assessing them. In real-world case studies where location choices of agents depend on numerous variables, the number of feedback effects will be increased accordingly. The mechanisms driving the sensitivity of the outputs of LUTI models to spatial biases are thus likely to be obfuscated.

5.5 Polycentric case study

5.5.1 Methodological choices

This experiment aims at assessing more in-depth the influence of the *boundary effect* on the outputs of LUTI models. We consider the case of a polycentric metropolitan area: the studied theoretical urban environment is composed of the "catchment areas" of two CBDs (called here East and West) separated by a suburban area (Figure 5.10). This study area consists in a rectangle of 60 by 25 kilometres. Since it is larger than the mono centric case study, the BSUs are reduced to squares of 1 by 1 km. Each CBD is located in the centre of a catchment area of 25 by 25 zones. The initial number of non-home-based jobs and households per zone is a function of the Euclidean distance to both CBDs, as for the mono centric configuration (see section 5.3).

Main inputs are presented in Figure 5.11. The simulation period is limited here to 10 years, and a linear growth of 1% per annum for both households and jobs is assumed. As in the mono centric configuration (section 5.4) each iteration of *UrbanSim* accounts for one year. A *MATsim* run is performed with an interval of three iterations. Three situations are defined in terms of size: equal-sized CBDs, small West CBD (West CBD half the size of the East CBD for households and jobs) and large West CBD (West CBD twice East CBD). Seven different extensions of the study area are considered: the *Complete* area and 6 *Small Extents* (named *Exx*, with *xx* the number of columns from the western extremity - see Figure 5.10). They result in a progressive inclusion of the East CBD into the studied area (see Figure 5.11, in appendix). Note that it means that the small west CBD and large west CBD case studies are not symmetric. For the former, the *small extents* result in the progressive inclusion of a larger CBD than the west one, while for the latter this is the opposite. Each pair of CBD's size and Extent is simulated 30 times to cope with the stochastic nature of the model; the results presented here are average values.

Each *Extent* is a subset of the *Complete* area. Hence, initial conditions of a zone are the same for all subdivision of the study area (e.g. same number of agents and level of real estate prices in t_0), at two exceptions: (1) the Euclidean distance to the CBD is computed to the closest CBD for the *Complete* area, and to the West CBD for all others. (2) The travel model (see section 5.3) estimates the car accessibility to jobs independently for each *extent*.

5.5.2 Estimation of the econometric sub models

The location choice sub model for households relies on three variables: car accessibility to jobs, residential buildings' real estate prices and Euclidean distance to the closest CBD. The latter one allows having the expected sign for parameter estimates of car accessibility and real estate prices (i.e. positive and

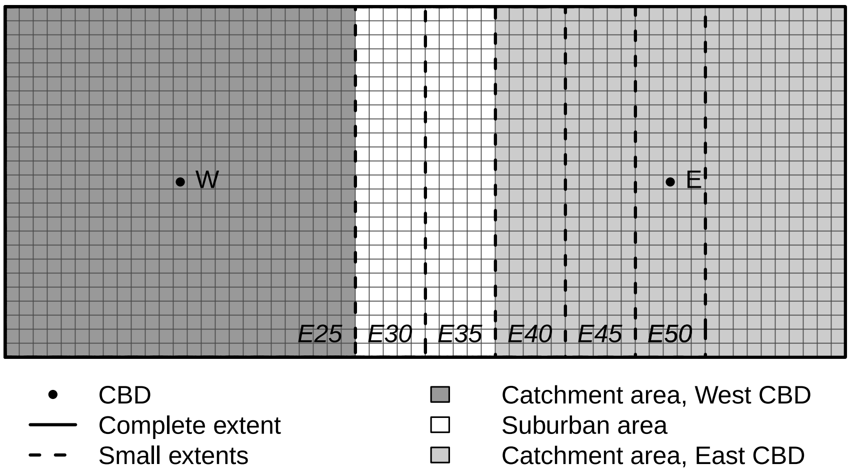

Figure 5.10 – Poly centric case studies (with *Exx* the name of the *small extents*, where *xx* refers to the number of column from the western extremity)

negative) while keeping the model as simple as possible (the distance to CBD is constant over time). Non-home-based jobs only depend on the car accessibility to jobs and on real estate prices of non-residential buildings. The main reasons for the selection of these variables is that they are grounded in the economic geography theory and are updated at each iteration (see section 5.2).

Estimations show a very low predictive power for households (Table D.4) and further specifications (not reported) did not allow solving this issue. The goodness-of-fit of the employment sub model is high in all cases (Table D.5). Note that parameter estimates for real estate prices are positive in the employment' sub model for most *Extents*. It means that the direction of the feedback loop will be opposite to the expected one (positive for car accessibility, negative for real estate price), therefore affecting the behaviour of the model. Despite these specificities, high utilities are found close to the CBD(s) in all cases, and future or relocating agents should thus locate there rather than in peripheral areas.

For the real estate price sub model, selected independent factors are the population and jobs densities (Table D.6). As expected, all parameter estimates are positive. Three additional econometric sub models intervene in one iteration sequence: (1) the home-based jobs location choice sub model forecasts

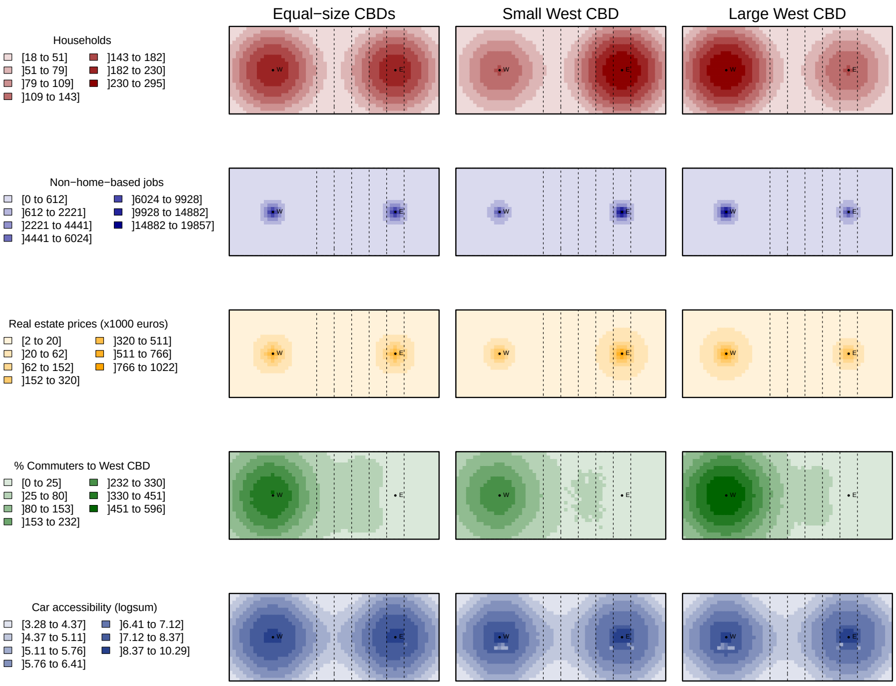

Figure 5.11 – Main characteristics of the poly centric case studies (discretisation: jenks)

the location choice of home-based jobs, assumed here to depend only on population density (negative parameter estimate in all cases), (2) residential and (3) non-residential location choice sub models determine the location of the new real estate projects, increasing the capacity of each zone to accommodate new or relocating households/jobs. Both depend on real estate prices and on car accessibility. The expected sign of the parameter estimates (i.e. negative and positive) is obtained in all cases.

5.5.3 Results

Since agents are homogenous, a limited number of indicators can be used to compare the final situations forecasted by the model. Four have been selected: the final number of (1) households and (2) non-home-based jobs per zone in t_{10}, (3) the evolution (between t_0 and t_{10}) of real estate prices for non-residential buildings and (4) the mean home-to-work travel time in t_{10}. The *Complete Extent* is used as reference, and relative differences (in %) are computed with

Small Extents: a negative difference means larger final values for the *Small Extent* than for the *Complete*, and vice-versa. Deviations from the *Complete Extent* are first computed for each BSU. Such zone-to-zone variations reveal a certain level of noises (due to stochastic variations in the *UrbanSim*), which vary upon the indicators. Hence, spatially aggregated variations have also been computed for three macro-zones: the "Catchment area of the West CBD", the "Suburban area", and the "Catchment area of the East CBD".

Let us start by examining the evolution of the *Complete* study area between t_0 and t_{10}. Table 5.6 shows that the magnitude of the values per zone in t_{10} remains credible. The largest differences with t_0 are observed for employment and can be explained by the low number of jobs in many peripheral zones (Figure 5.11). Moreover, most of the variations take place near the CBDs, a result consistent with the parameter estimates of econometric sub models (see Table D.4 to D.6). Hence, the behaviour of the model appears believable, and its outputs can thus be used to assess the sensitivity of LUTI models to the delineation of the study area.

Indicator	**Value in t_{10}**			ρ
	Min	Mean	Max	
Households	21.93	110.40	243.97	-0.97
Jobs	0	211.49	16 371	-0.38
Real estate prices	2 109	21 382	820 021	-0.47
Home-to-work travel time	0	60.95	119.19	0.86
	Variations with t_0 (%)			
Households	9.44	10.56	14.13	-0.97
Jobs	-100	31.45	100	-0.15
Real estate prices	3.59	6.75	10.48	-0.47

Table 5.6 – **Calibration** (*Complete* extent and equal-size CBDs; ρ = pearson product-moment correlation with the Euclidean distance to the CBD; all coefficients significant for $\alpha \leq 0,001$)

Households

Variations (between *Complete* and *Small Extents*) are limited in magnitude, either by zone (Figure 5.12) or by macro-zones (Table D.7). Since their magnitude is lower than the inter-runs variations (which vary from 0.58 to 1.27%), they can only be considered as random; indeed no specific spatial structure appears. The size of the CBDs seems to slightly influence these variations (that decrease from Small West CBD to Large West CBD), but it could be an artefact due to the higher number of agents. Note that the growth in the number of households between t_0 and t_{10} is larger close to the CBDs, and that this structure is preserved in all extents of the study area.

Non-home-based jobs

Large differences are observed between the *Complete Extent* and *Small* ones, and they are significantly larger than inter-run variations. A strong spatial structure emerges in zone-to-zone (Figure 5.13) as well as in macro-zones variations (Table D.7) with, for all *Small Extents* and all CBDs' sizes a higher concentration of (non-home-based) jobs near the CBD (negative differences with the *Complete Extent*) and a lower one in the "dense suburbs". Zone-to-zone variations are more influenced by the extent than by the size of the CBDs. From Extent E35 to E25, the "dense suburbs" shrink to a ring of limited width. Zone-to-zone variations are low in the case of a Small West CBD, medium for the Large West CBD, and high for Equal-size CBDs. For macro-zones, negative differences are observed in the smaller CBDs, i.e. the portion of the East CBD for the Equal-size CBDs case study, and respectively the West and East CBD in the two other cases. Note that for E50 (Equal-size CBDs) and E40 (Large West CBD), the opposite situation appears. Due to the concentration of jobs close to the CBDs, no variation is observed for extents that do not include any part of the catchment area of the East CBD. Overall, between *Complete* and *Small Extents*, a relocation process is observed from the "suburbs" to the city centre. The intensity of this process highly depends on the size of the CBD, which can be explained by the larger number of new jobs. Across Extents, the magnitude remains similar. Four particular situations will be later discussed: E40 and E45, for Small and Large West CBD.

Real estate prices

Prices of non-residential buildings depend on population and jobs densities. The spatial structure of their variation between Extents is quite similar to that observed for jobs, but with a larger level of noise induced by the random variations observed for households. Moreover, the parameter estimates of the real estate price sub-model vary from one case study to another, therefore, affecting the evolution of these prices. Zone-to-zone variations show in most cases positive differences close to the CBD, meaning that the increase in real estate prices is lower in these zones for *Small Extents* than for the *Complete Extent* (Figure 5.14). Aggregation by macro-zones leads to the same conclusions (Table D.7).

As for jobs, E40 and E45 show the opposite situation for Small and Large West CBDs than for the *Complete Extent*, i.e. a larger increase of real estate prices close to the CBD (positive differences on Figure 5.14). A potential explanation lies in the parameter estimates of the (log of) population density in the real estate prices sub-model that is slightly larger for E40 and E45 than for the *Complete* study area. Hence, a comparable increase of the population in this zone (Figure 5.12) may lead to these counter-intuitive results.

Finally, parameter estimates of the (log of) job density are lower for the *Complete* than for most *Small Extents* (Table D.7). Hence, an equal increase in the number of jobs per zone will produce a larger increase of real estate prices in the *Complete Extent*, which reinforces the positive differences observed.

Home-to-work travel time

The magnitude of the variations is larger than those observed for the previous indicators, and the spatial structure is clear and highly influenced by the relative size of the West CBD (Figure 5.15). The observed noises are probably explained by the use of only 25% of the agents in *MATsim*'s runs.

A large increase in commuting time is observed for all zones located East of the West CBD in the Small West CBD case; the same appears in E50 for the Equal-sized CBDs and Large West CBD cases, but it is limited to the catchment area of the East CBD. On the contrary, larger commuting times are observed in the Eastern part of the study area for E35 to E45 (Equal-size CBDs and Large west CBD case studies). The variations aggregated by macro-zones (Table D.7) show a large decrease in home-to-work travel time for all Extents in the Small West CBD case and for E25 to E35 in the equal-size CBDs case. Meanwhile, an increase in commuting times is observed for the catchment area of the East CBD for E40 and E45 in the Equal-size CBDs and Large West CBD cases. This can easily be explained by (1) the lower overall competition for jobs in *Small Extents* (the number of households decreases more rapidly than the number of jobs when the size of the extent decreases, see Figure 5.11). (2) The exclusion of the East CBD from the study area leads to lower opportunities for people located in the eastern part of the study area, constrained to commute to the West CBD in all extents smaller than E50. The same is true for the western part of the study area where residents face the competition of people that were commuting to the East CBD in the *Complete extent.* The combination of these factors induces the observed decrease of average home-to-work travel time, and their local increase.

5.5.4 Discussion

Two main critics can be addressed to our methodology. First, the different extents of the study area were not designed with a concern of realism, but to provide a continuous evaluation of the influence of cities' delineations. However, we end up with quite realistic situations such as Extents E25 to E35 that consist in adding to the study CBD a rural area with few or no functional links with the CBD. On the opposite, Extents E40 to E50 include a portion of another CBD into the studied urban area. Secondly, the study area is here totally

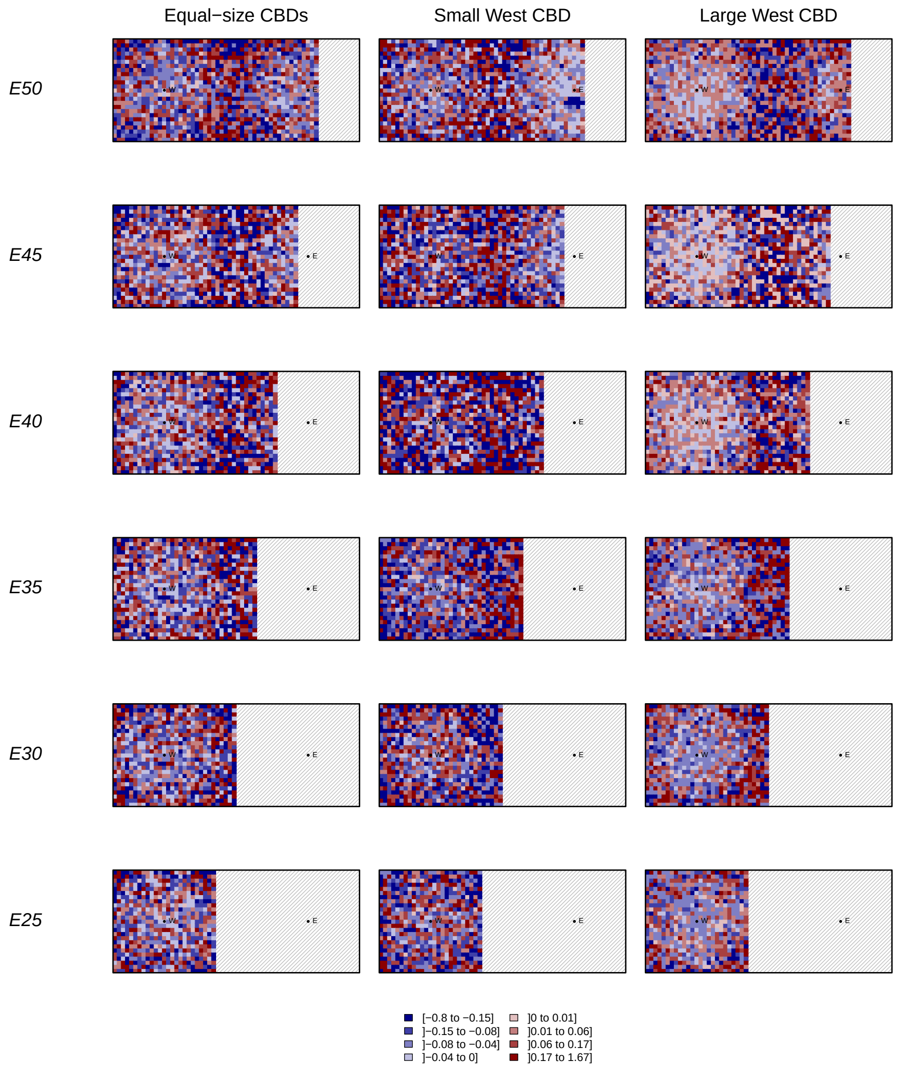

Figure 5.12 – Variations of the number of households (value = $\frac{Complete - E_x}{Complete}$, in %; discretisation method = quantiles; dashed grey = not included in the extent)

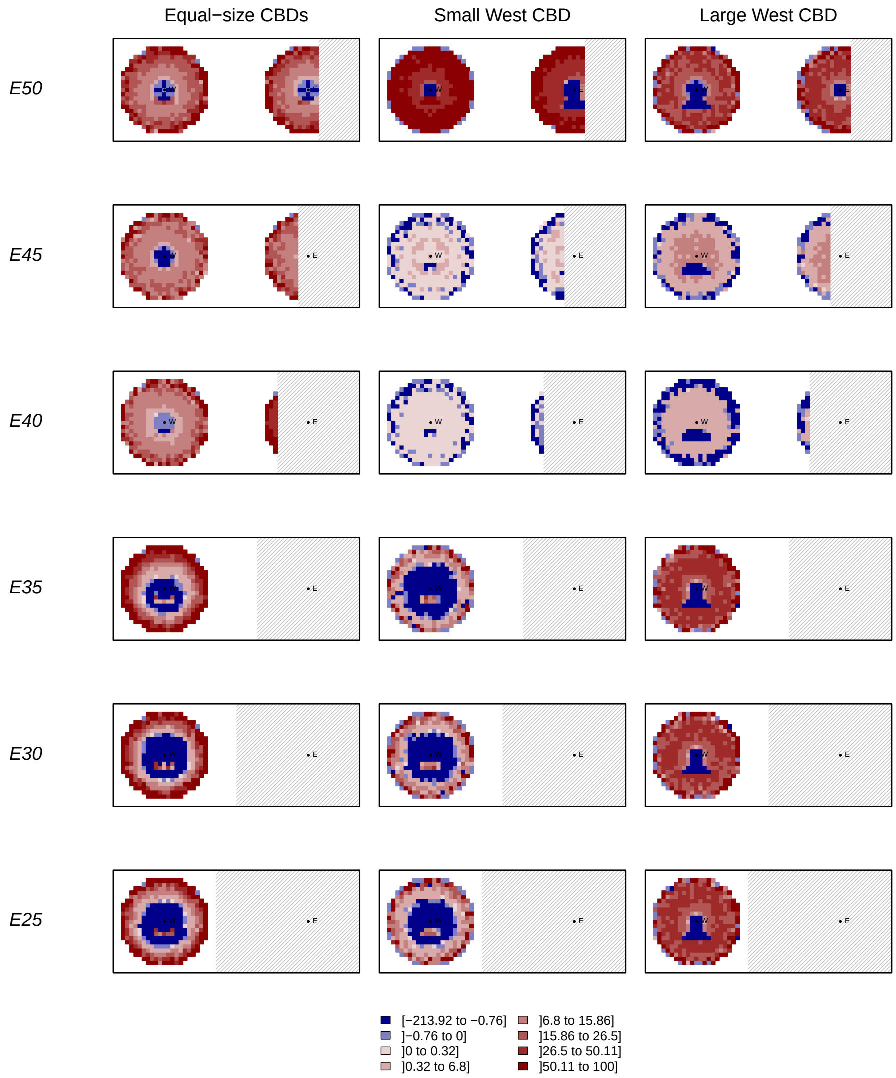

Figure 5.13 – Variations of the number of jobs (value $= \frac{Complete - E_x}{Complete}$, in %; discretisation method = quantiles; dashed grey = not included in the extent; white = no jobs in t_0)

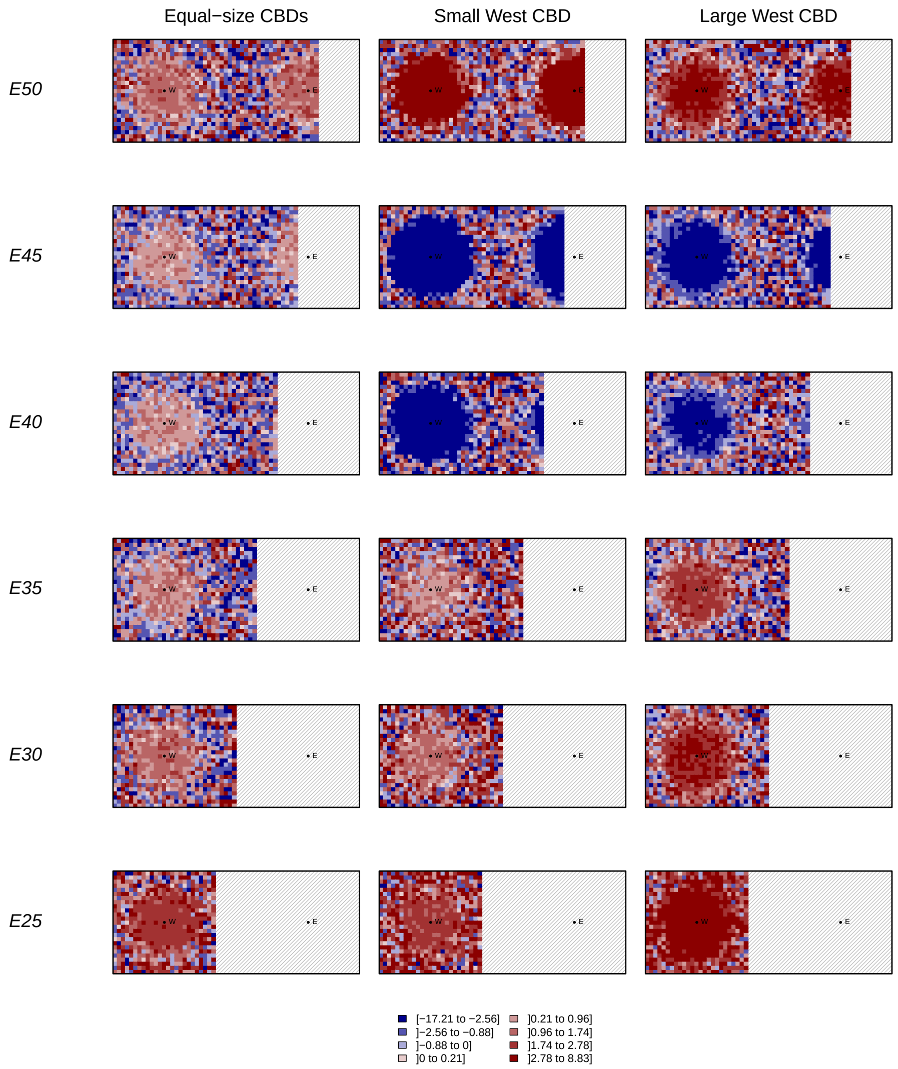

Figure 5.14 – Variations of real estate prices (non residential prices in t_{10} minus t_0; value $= \frac{Complete - E_x}{Complete}$, in %; discretisation method = quantiles; dashed grey = not included in the extent)

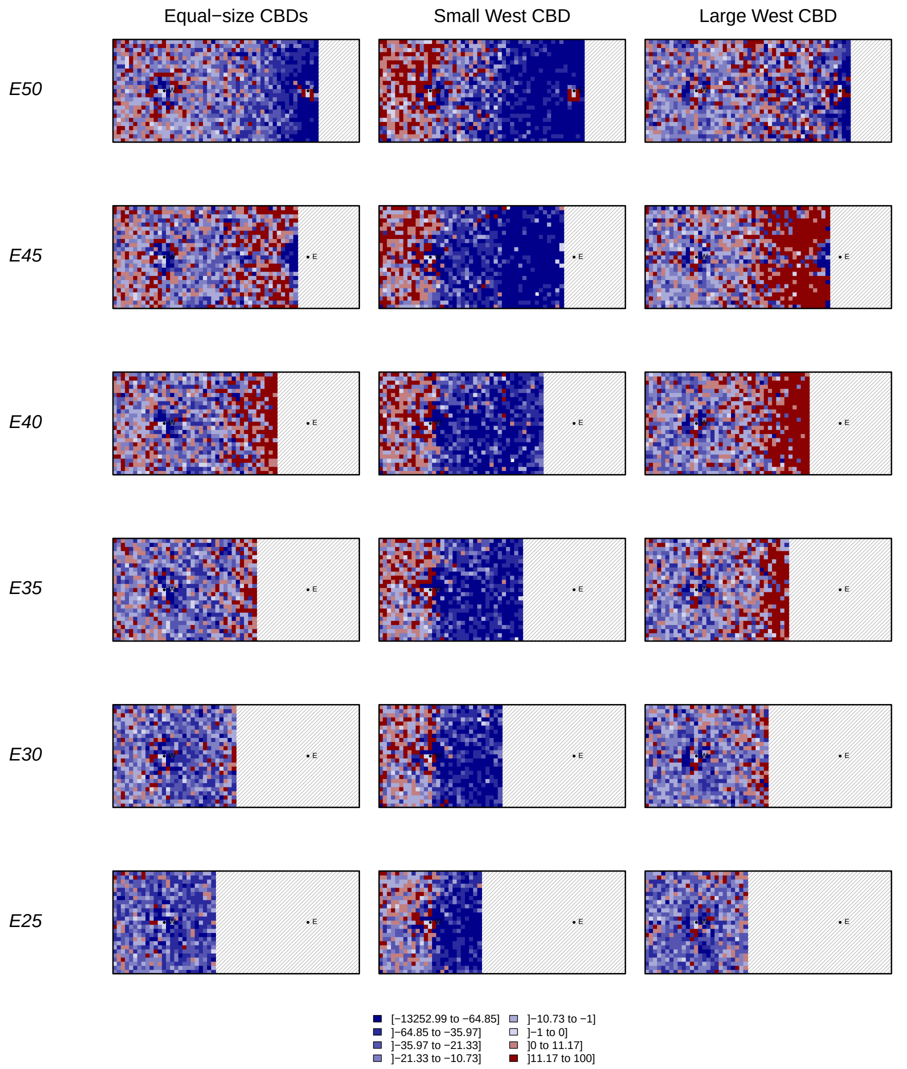

Figure 5.15 – Variations of the home-to-work travel time (value = $\frac{Complete - E_x}{Complete}$, in %; discretisation method = quantiles; dashed grey = not included in the extent)

isolated from the "rest of the world". The *UrbanSim* model indeed assumes a closed study area while our results suggest that taking into account fluxes from and to the "rest of the world" would have been preferable, especially in terms of home-to-work travel. Such strategies are often implemented into the transport model of LUTIs' applications (see e.g. Gerber, 2012; Gerber et al., 2012).

Nevertheless, most spatial structures and parameter estimates of the econometric values in the sub-models are consistent with the principles of the Alonso-Muth model. Hence, the evolution of the study area over the simulation period is credible (Table 5.6). The variations observed in primary (location of households and jobs) and secondary indicators (evolution of real estate prices, home-to-work travel time) are those expected.

In any case, our results suggest that the addition of rural areas has only a limited influence on the model. It corresponds to the E30 delineation, where the relative variations with the *Complete Extent* mostly consist in noise. Strong biases arise when the study area (1) fails to encompass some portions of the influence area of one CBD (Extent E50, where peripheral areas of the East CBD are excluded) or (2) includes some portions of the catchment area of a close CBD, but leaves this latter centre not included (Extents E35 to E45, that does not include the East CBD itself). These cases show major discrepancies from the *Complete Extent*, that may lead, in operational applications of LUTI models, to falsehood in transport or land use planning scenario's evaluations. Moreover, Extent E50 indicates that the magnitude of these biases is proportional to the size of the not included CBD (i.e. biases increase when the East CBD is larger than the one on which the study area is centred (West CBD).

5.6 Implications for real-world applications

This chapter relies on a synthetic case study. It was designed to simplify the structure of the city and, therefore, exploring more easily how spatial biases affect the behaviour of the model. We faced, however, several unexpected difficulties in the development and applications of this synthetic case study. The final mono centric configuration was mostly determined by trials and errors. Previous simulations on a smaller grid (27 by 27) were unreliable, due to the small number of aggregated BSUs. Moreover, relying on the Euclidean distance to the CBD produces a concentric extension of the city, which does not match the boundaries of a squared study area. Peculiarities were thus observed on the edge of the grid; hence the decision to reduce the mono centric configuration to a circular study area. The location choice sub model also raises two difficulties: first, a low goodness-of-fit is observed for households, even if the explanatory variables are highly correlated with the population density. Secondly, parameter estimates often have a counter intuitive sign, which cannot be explained by multi collinearity issues. Overall, the goal followed has been partially reached, as showed by section 5.4.3 and 5.5.4. The usefulness of the synthetic case study is, however, reduced by the difficulty to calibrate the location choice sub models, even on a city with a perfectly known spatial structure.

Some implications can, nevertheless, be drawn for real-world applications of LUTI models. First, the analysis of the feedback effects affecting the behaviour of the model (section 5.4.3) raises the question of explanatory versus predictive models (see Shmueli, 2011). That is to say, should the location choices sub model reflect the process that are supposed, according to literature, to affect households' location choices, or maximise the goodness-of-fit even if the included variables have no clear meaning on an economic or geographical point of view? Nguyen-Luong (2008) argues, using the case study of Paris, that simple specifications with few variables are preferable (although the aim of this work is obviously different than ours). This approach is also privileged here. There is, nevertheless, a direct link between the variables included in these sub models and the behaviour of the model. The non inclusion of the car accessibility to jobs, for instance, will neutralise the potential influence of an improvement of transport infrastructure (as in the *Highway* scenario). Hence, the variables used in econometric sub models of *UrbanSim* define the utility level of the agents and the feedback effects accounted for. The issue here is that the different sub models are often estimated separately, leading to counter factual situations (see chapter 7).

Our results are non conclusive for the *scale effect* (which will be studied more in-depth in chapter 6), due to the peculiarities observed in the behaviour of the *Aggregated* case study (section 5.4.3). Simulations clearly show, on the contrary, that the *boundary effect* (i.e. the delineation of the study area) should be of primary concern in operational applications of LUTI models. Systematic variations are observed across extents for most indicators. They do not consist in stochastic noises, but have a strong spatial structure that can be explained by the nature of the study area (inclusion or exclusion of the eastern CBD, see section 5.5.3). This is both good and bad news. On the plus side, it means that a careful exploratory spatial data analysis for defining the relevance of the study area can clearly improve the quality of LUTI models' outcomes. The bad news are that the study area in often imposed by the authority or the sponsor of the project (i.e. an administrative authority will most likely require its entire area of jurisdiction to be included in the model, even if such delineation is not meaningful in terms of urban reality). The study area should thus be carefully delineated to avoid or to control the inclusion of a portion of the influence area of other cities. Chapter 7 proposes an "ideal" method of cities delineation for LUTI models applications. The comparability of LUTI model outputs is, in any case, likely to remain difficult when study areas are of different nature. The choices made in the delineation of the study area should, therefore, be made very clear in the description of any operational application of LUTI models.

CHAPTER 6

The Brussels case study: lessons for policy evaluations

6.1 Introduction

Ou final experiment is to assess if spatial biases may jeopardise policy evaluations based on LUTI models' outputs. We know from chapter 5 that these outputs vary due to both the *scale* and *boundary effects*. The analyses conducted in this previous chapter were, however, limited to a baseline situation, i.e. an uneventful evolution of the study area. The two scenarios implemented in chapter 5 have been used to provide a reference, for comparing the variations observed for different *spatial extent* and *spatial resolution*. We intend to go one step further in the present chapter, by focusing on potential wrongheadedness induced by spatial biases in policies' evaluation.

That is to say, can the implementation of a given land-use or transport scenario be considered profitable when the model is run at one BSU level, but unprofitable on another one? We rely here on the urban region of Brussels (Belgium) as empirical case study, for reasons exposed in chapter 1. For consistency purpose, we also limit ourselves here to the *scale effect*.

Five different scenarios are defined, covering various dimensions of the model system. The outputs of the *UrbanSim* model for four different BSU levels are used to compute sustainability indicators, and to perform a generalised-costs comparison of the scenarios. The model system is thus considered here as a "black box". To answer the research question, the variations of the indicators across scenarios are compared to those between BSU levels (for both their magnitude and direction). The generalised-costs approach is used to assess if the ranking of each scenario varies from one BSU level to another. Figure 6.1 details this workflow.

The chapter is organised as follows. Section 6.2 shortly reviews the methods used in conjunction with LUTI for policy evaluations. Data and methodology are detailed in section 6.3. Section 6.4 presents the results that are discussed in section 6.5. Section 6.6 concludes.

6.2 Policy evaluation in LUTI models

To select the indicators that will be used for comparing the scenarios, it is necessary to summarise how policy evaluation is performed in land use and transport planning, and the specificities of LUTI models. This evaluation usually relies on multi-criteria analysis or cost-benefits analysis (Geurs and van Wee, 2004; van Wee, 2015). Both methods are widely described throughout the literature. The reader can refer to e.g. Ishizaka and Nemry (2013), Nijkamp and Blaas (1994) for the former one, and to Atkinson and Mourato (2006), Boardman et al. (2006) for the latter. Nevertheless, different conceptual and technical challenges restrain the capabilities of LUTI models to perform policy evaluations.

The main criterion to evaluate a policy appears to be the sustainability of the city. This notion is not exactly the same from economic (see e.g. Arrow et al., 2004) or a geographic point of views (see e.g. Brown et al., 1987; Bulkeley and Betsill, 2005). In an urban/LUTI model perspective, the principal component is the influence of transport on environment (Geurs and van Wee, 2004; Rodrigue et al., 2009). The main conceptual challenge is thus to make this influence endogenous in the model. Car ownership and air pollution level, for instance, are rarely estimated (Wegener, 2004; Dowling, 2005). Therefore, without additional methods or model, sustainability indicators estimated from the outputs of LUTI models are limited, and often rely on expert judgment (Geurs and van Wee, 2004).

Policy evaluation (i.e. the comparison between different scenarios) is, by definition, the last step of LUTI model projects. As noted by Wegener (2011a), many operational applications have run out of time before reaching this stage due to unexpected practical difficulties. Moreover, despite the large variety of detailed data required, one common and ancient (see Lee, 1973) criticism of LUTI model is that they produce only aggregated results. Again, sustainability indicators seem particularly prone to be available only at a meso- or study area level (Geurs and van Wee, 2004). Efthymiou et al. (2014) proposed multidimensional indicators to take advantage of the disaggregated nature (in space, time, and agents) of LUTI models. This approach has, however, not yet been used in operational applications.

Eventually, the main difficulty is the multidimensional nature of the sustainability, encompassing economic, social, and environmental components (Hély and Antoni, 2014; Proost et al., 2015). Aggregation can be done qualitatively, but a formal integration of these three pillars requires each indicator to be expressed in monetary units. For that purpose, Proost et al. (2015) have proposed a Social Welfare (SW) function that allows comparing the outcome of one scenario to those of the baseline. Its specification consists in a weighted sum of the utility of (a) the inhabitants, (b) the commuters, and (c) the rest of the world. The (d) local stock left to future generations is also included (accounting for the sustainability), together with (e) the cost of implementation of the scenario and (f) the generated revenues. Using the case studies of the EU-funded *SustainCity* project (Brussels, Paris, and Zürich) as examples, Proost et al. (2015) conclude, however, that the translation of this SW function from a theoretical formulation to a practical indicator is severely limited by data availability (see section 6.4).

To sum up, in operational applications, policy evaluation is often limited to a set of simple indicators (see Bartholomew, 2007 for review). Hence, the same approach will first be used in this chapter. These simple indicators of the influence of the scenarios are assumed to have an equal weight and will, therefore, be examined independently rather than in a multi-criteria analysis. In a second step, the SW level of each scenario will be computed, in order to rank them according to their profitability. The results of both methods will be compared for our different BSU levels.

6.3 Data and methodology

6.3.1 Data

The empirical case study used in this chapter is the metropolitan area of Brussels (Belgium), based on the Brussels case study of *SustainCity*. A detailed description of this model and the database can be found in Cabrita et al. (2015). Note that Patterson et al. (2010), Patterson and Bierlaire (2010) described earlier implementations of *UrbanSim* on Brussels. However, their works consist in prototype models, based on aggregated data, and have few common points with the model of Cabrita et al. (2015).

The study area used in the *SustainCity* project (or "SustainCity Area") raises several concerns (see Thomas et al., 2015; and chapter 3), since it encompasses municipalities having few relationships with the CBD and/or that belongs to the catchment area of another city (see Figure B.2). Chapter 5 have demonstrated that the extent and composition of the study area may influence

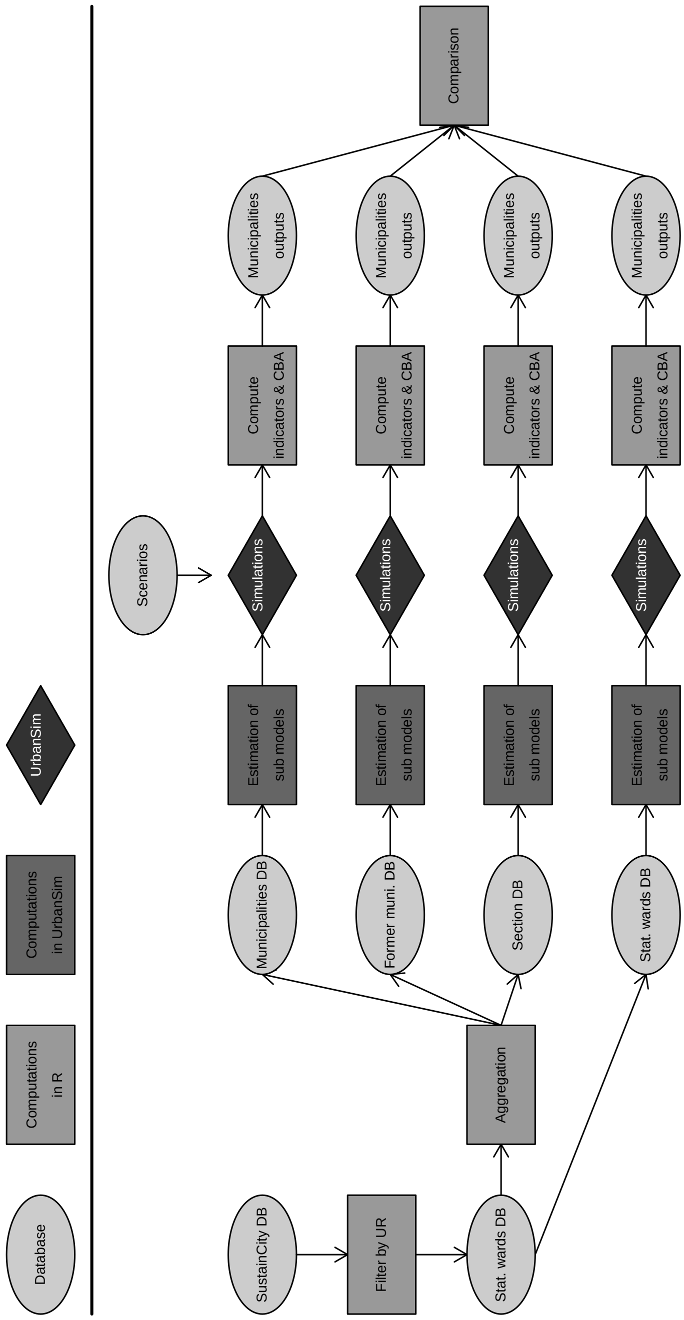

Figure 6.1 – Workflow

BSU level	n	Surface (km^2) Min	Mean	Max	Inhabitants Min	Mean	Max
Statistical ward	2 074	0.01	0.74	13.70	0	694	4 608
Section	550	0.01	2.78	15.47	0	2 616	18 883
Former muni.	173	0.08	8.82	45.02	0	8 316	77 238
Municipalities	62	1.16	24.62	68.58	3 282	23 205	104 698

Table 6.1 – **BSU levels** (descriptive statistics for 2001; study area = urban region of Brussels)

the parameter estimates of the location choice's sub models within *UrbanSim* and the outcomes from the model.

It has thus been decided to reduce the "SustainCity Area" to a more meaningful delineation of Brussels, the Urban Region (or UR, see Van Hecke et al., 2009 and Figure 6.2). The UR has an extent of 1 526 km^2 and in 2001(base year of the model) it accounted for 1.44 million inhabitants and 0.99 million jobs (see also Figure 6.4). Its centre is composed of the 19 municipalities of the Brussels-Capital Region, hereafter BCR. For comparison, the "SustainCity area" reaches 5 169 km^2, representing 2.69 million of inhabitants and 1.45 million jobs.

Cabrita et al. (2015) uses *Statistical wards* as BSUs. These are the smallest areal units for which statistical data are available from the Belgian Directorate General Statistics and Economic Information (DGSIE). They can be aggregated into larger nested BSU levels: *Sections*, *Former Municipalities*, and *Municipalities*. They will al be used in the simulations. Table 6.1 summarises their relative size.

6.3.2 Methodology

The model

The *UrbanSim* model is used to forecast the evolution of the Urban Region of Brussels. A detailed description of the model system can be found in chapter 5. The base-year data is 2001 (allowing the database to rely on the 2001 population census). Since forecasting the evolution of the city far into the future is not mandatory here, the simulation period is limited to 20 years (i.e. 20 iterations of *UrbanSim*). *MATsim* is run with an interval of five iterations, i.e. in 2001, 2006, 2011, and 2016. Calibration is performed after ten iterations (i.e. in 2011, year of the last existing population census). Simulations are performed independently for the four BSU levels. Moreover, five scenarios are defined, and it is assumed that they are implemented at the end of 2010. Note that the simultaneous implementation of different scenarios is not considered.

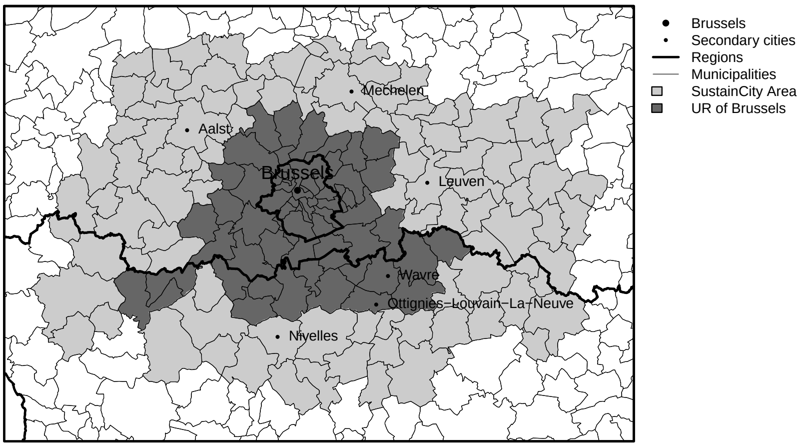

Figure 6.2 – **Study area** (Urban Region of Brussels, relative to (a) the "SustainCity Area" and (b) to Belgium; typology of the municipalities from Van Hecke et al., 2009)

Hence, we end up with 20 combinations of BSU level and scenario. To cope with the partially stochastic nature of *UrbanSim*, each of them is simulated 10 times. In section 6.4, all indicators are based on average values.

Scenarios

To provide a reference, a business-as-usual situation is first defined. In this Baseline, the simulation period is uneventful. The only variations are the growth of the number of households and jobs. These forecasts come from the Belgian Federal Planning Bureau and are detailed in Cabrita et al. (2015). For households, it is of about 1% per annum, from 743 487 in 2001 to 887 138 in 2020. Growth rates vary widely across employment sectors, from 9% (hotel/restaurant) to 59% (leisure activities). In absolute term, the largest increases are observed for tertiary sector' jobs (+86 250), health (+44 766), and industrial activities (+32 320). Note that the variations of the number of agents over the simulation period are assumed to be equal for all scenarios.

A *Cordon* toll scenario is first designed, to reduce the congestion problems to or within the BCR. Hence, the toll barrier is located on the boundaries of the BCR (Figure 6.3). The fee is fixed to 5€, as in Cabrita et al. (2015), and has to be paid whatever the time of the day (this choice is the result of modelling constraint, see section 6.5). Practical implementation rely on MATsim, using an external configuration file defining the links of the road network affected, the type of congestion tax, the monetary value of the fee, and the period of time. For details see Cabrita et al. (2015) and Nicolai and Nagel (2015).

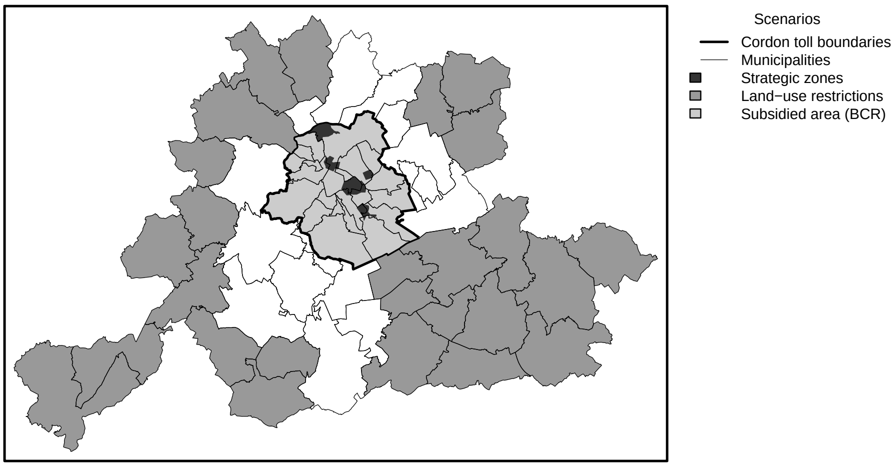

Figure 6.3 – **Scenarios** (map of their implementation area)

The *Office* scenario builds on the 11 strategic zones identified by the BCR' territorial development agency for future urban development (see ADT-ATO, 2015). A review of impact studies available from the ADT-ATO website indicates that new office spaces are planned for five of them (Figure 6.3), accounting for a total of about 440 000 square meters of floor space. We assume that these development projects take place at the level of the statistical wards and are later aggregated into larger BSU levels. The implementation of this scenario relies on the Scheduled Development Event Model (SDEM) within *UrbanSim* (see Gallay, 2010). Practically speaking, this sub model will increase the values of non-residential square meters available in each zone by the required amount, in 2010.

Finally, the *Subsidy* and *Land-use* scenarios aim at reducing urban sprawl of households. For the first one, we assume that a fiscal incentive is allocated to households choosing to locate within the BCR. It accounts for an apparent decrease of real estate prices by 5%. The Land-use scenarios introduce stronger land-use planning regulations, decreasing the number of future residential units allowed in the suburbs of Brussels by 20%. Suburbs (see Figure 6.3) are defined here accordingly to Van Hecke et al. (2009). Both scenarios are implemented in *UrbanSim* by altering the average value per dwelling (*Subsidy*) or the residential units capacity (*Land-use*) within each zone, by means of the SDEM sub model.

Estimation of ***UrbanSim*** econometric submodels

The last step prior to running the simulations is to estimate the econometric sub models within *UrbanSim* (see chapter 3, chapter 4, and chapter 5 for a complete description). The aim of this chapter calls for a workflow where these sub models are estimated independently for all BSU levels (Figure 6.1). Hence, to ensure consistency through BSU levels and minimise potential selection bias, we rely on automatic variable's selection procedures.

Update of real estate prices at the end of the iteration is performed in *UrbanSim* by a log-linear regression estimated by OLS (see chapter 3). This specification was replicated in R and then calibrated by a backward procedure (iterative exclusion of independent variables whose t-test' significance level is higher than 0.05, starting by the least significant one). The main weakness of this approach, as shown in Table E.3, is that several specifications are limited to fixed factors and/or to only one endogenous variable. Therefore, variations of population or jobs' densities will not influence these prices. Although this situation is unsatisfactory from a modelling point of view, we have decided to keep these specifications, to ensure consistency in the estimations.

The econometric framework of the location choice sub models is detailed in chapter 4. They forecast the probability of a given building to be selected by new or relocating agents (households and jobs), or for real estate development project. Calibration is handled by estimating, within *UrbanSim*, ten specifications based on different combinations of the explanatory factors (Table E.1). The specification having the lowest AIC value is selected. Note that for employment, estimations are only performed for non-home-based jobs. Given the lack of accurate data on home-based jobs, it has been decided to neutralise that part of the model system.

It should be noted that for *Former municipalities* and *Municipalities*, this procedure leads to frequent inclusion of non-significant variables (especially for employment, see Table E.5). Hence, the AIC was perhaps not the best indicator (see e.g. Burnham and Anderson, 2002). However, the sub models have been estimated within *UrbanSim*, in which no other indicators were available for model comparison. To ensure the reproducibility of the results, it has been decided to stick to these specifications.

Endogenous variables (i.e. updated by *UrbanSim* during its iterations) are used as much as possible. Three constant variables are nevertheless considered to account for characteristics that cannot be forecasted by the model (Table E.1). The following indicators account for agglomeration economies: Population density is defined as the number of inhabitants per square kilometre in each BSU. Job density (in jobs per sq. km) is declined for the total jobs and for each of the eight activities sectors. Accessibility factors are the Car Accessibility to Jobs, a logsum indicator estimated by *MATsim* (see Nicolai and Nagel, 2011 for details), and the Euclidean distance (in meters) to the Brussels'

CBD. This latter variable is measured between the centroid of the municipality of Brussels and the centroid of all other BSUs.

Socio-economic amenities include the share of households within a BSU having a monthly income higher than 3 100 €, or lower than 1 852 €. These categories come from the 2001 population census. In the households' location choice sub model (or HLCM), these variables are used as interaction term with a dummy variable equal to one if the household has a high income, and zero otherwise. The percentage of households within a BSU where (at least) one member has a university degree is also considered.

Other amenities are the House prices, the level of local taxes, and a Green amenities score. Available data on real estate prices are highly limited in Belgium (see chapter 3). House prices serves thus as a proxy for all residential real estate types, and identical values are imputed to all BSUs belonging to the same municipality. Local taxes is an instrumental variable, included only into the real estate price' sub model (or REPM), to reduce potential endogeneity biases. Note that their level is limited to 9% of the federal taxes (i.e. if an household pays 10 000 €/year of taxes to the state, its municipality of residence can charge in local taxes a maximum of 900€). The Green Amenities score is computed from the surface of each BSU covered by green areas (forest or agricultural land, data from the CORINE 2006 Land Cover Database; see EEA, 2006) divided by the total surface.

Descriptive statistics at the level of the statistical wards' level are given in Table E.2. For larger BSU levels, their value is computed by aggregating the database by sum or means. Note that in the econometric sub models, most of these explanatory factors are expressed logarithmically. Figure 6.4 shows their spatial distribution.

6.3.3 Policy evaluation

Simple indicators are first computed, for each BSU level and each scenario. Agent's location choices are examined by computing the share of (1) households, (2) tertiary sector jobs, and (3) total jobs within the BCR at the end of the simulation period (i.e. 2020). At the study area' level, the home-to-work (1) travel times and (2) distances are used, together with (3) green space consumptions. By selecting these indicators, we attempt to cover different components of both the model systems and the sustainability issue. They also have the advantage of being direct outputs of either *UrbanSim* or *MATsim*. The variations in-between scenarios are later compared to those across BSU levels, to assess the sensitivity of the model to each component.

In the second step, a generalised-cost approach is used, by computing the Social Welfare (SW) level of each scenario (as proposed by Proost et al., 2015).

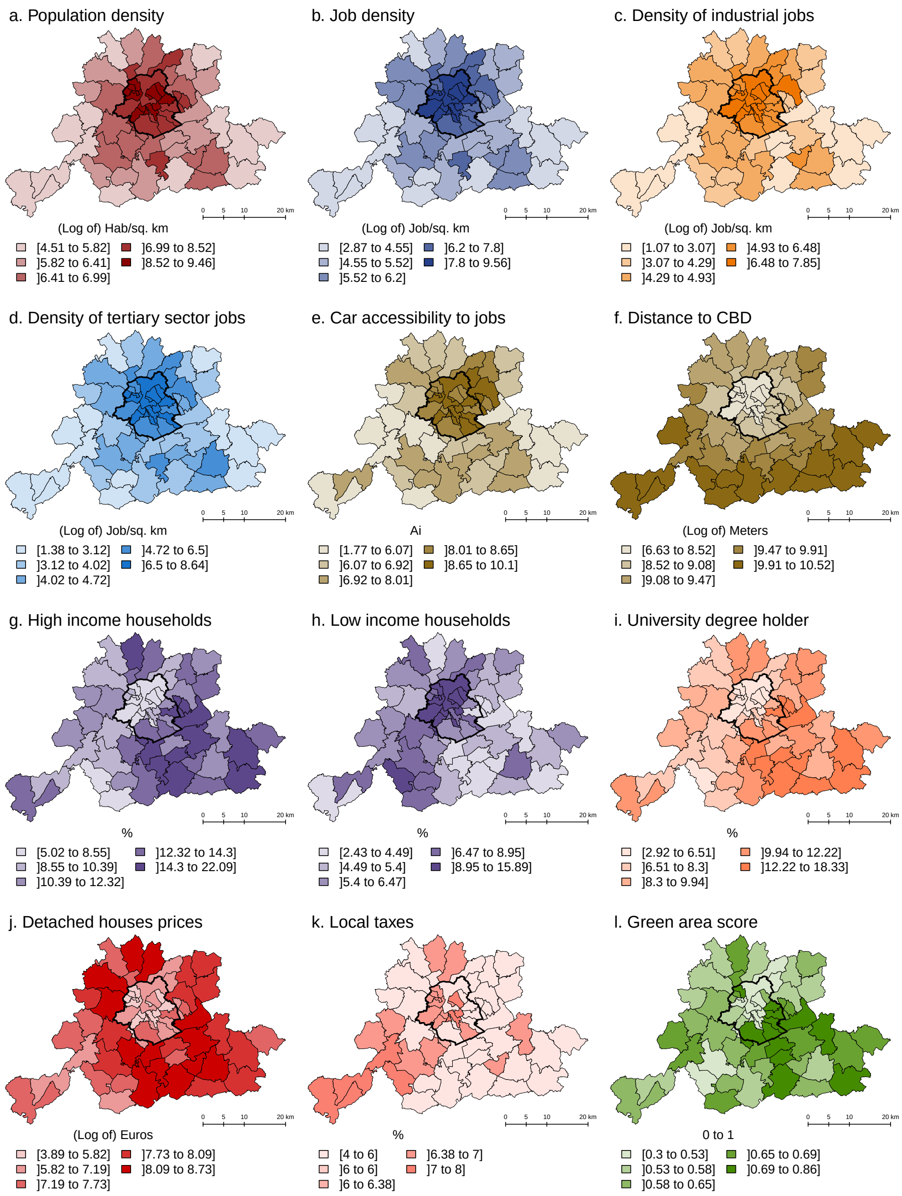

Figure 6.4 – Main variables (BSU level = municipalities)

The SW level (see equation 6.1), has the advantage (compared to the simple indicators) of allowing ranking the scenarios.

$$SW = U + (R - C) \tag{6.1}$$

In equation (6.1), U is the final utility level at the case study level (i.e. the sum of the utility of all inhabitants), R the direct revenues generated by the scenario, and C the implementation costs. Both R and C are computed at the public authorities' level. This specification comes from Proost et al. (2015). Nevertheless, several simplifying assumptions are made: (a) the city is closed, meaning that the utility levels of commuters and the rest of the world are not taken into account. (b) The only time horizon considered is the end of the simulation period. (c) No assumptions are made on equity preferences. Therefore, an equal weight is imputed to all income classes. (d) The utility level U is computed as in equation 6.2.

$$U = INC - HC - TC \tag{6.2}$$

INC denotes the annual income of the inhabitants (note that it differs from R, the revenues generated by the scenarios). HC is the annual housing cost equal (as in Di Pasquale and Wheaton, 1996) to the selling price times the interest rate (set here to 5%). TC is the annual transport cost, equal to the product of the commuting time with the number of working days per year (220 here) and the value of time (0.15 €, per minute). All components of U are thus expressed in Euros. Note that the utility is thus reduced here to a budget constraint. We follow here the specification and parameters' values of Proost et al. (2015). Housing and transport costs are endogenous to the model system. Evolution of incomes, on the contrary, is not modelled. The earnings of each household are thus constant over the simulation period. Cost and revenues of the scenarios will be estimated based on the literature (see section 6.4).

All scenarios are supposed to increase the sustainability of the study area. The simple indicators computed at the study area' level allows estimating which scenario produces the more sustainable final situation. This approach, however, has not been followed here since the aim of this work is to compare through scales the variations induced by the implementation of the scenarios, not to assess which one is optimal from that sustainability point of view. It should also be noted that due to the simplifying assumptions made in the estimation of the SW level, the generalised-costs approach do not include a sustainability component.

6.4 Results

6.4.1 Econometric sub models

Table E.3 presents the parameter estimates of the real estate prices' sub model. For houses, all parameters have the expected sign. The adjusted R^2 increases from small to large BSU levels, while the number of significant variables decreases. The explanation for these variations is that even if the prices are uniform within the municipality, the distribution of the independent variable is not. The goodness-of-fit is lower for flats, but parameter estimates remains mostly of the expected sign. For the households' location choice sub model, the best specification (using the AIC criterion) is identical for the four BSU levels, and corresponds to the inclusion of all explanatory factors (although the goodness-of-fit remains low). In the case of employment, the goodness-of-fit is generally high, but the specifications vary more widely between BSU levels. Parameter estimates are of the expected sign in most cases for both households (Table E.4) and jobs (Table E.5).

Sub models forecasting the location of future residential and non-residential developments have been constrained to a single specification, due to the lack of data on developer's behaviour. For residential buildings, it involves the (log of) population density, Euclidean distance to the CBD, housing prices, and the car accessibility to jobs. The specification for non-residential buildings relies on the (log of) population density, job density, house prices, and on the car accessibility to jobs. Both parameter estimates are of the expected sign and further details are, therefore, not included.

Parameter estimates of both of these econometric methods are known to be sensitive to the *scale effect* of the MAUP (i.e. to a change of the size of the BSU). As expected, such variations are observed here. The variations of parameter estimates will, however, not be further discussed since we focus here on changes in the outputs of *UrbanSim*. We refer to Fotheringham and Wong (1991); Arauzo-Carod and Antolín-Manjón (2004); and chapter 4 for work dedicated to the MAUP.

6.4.2 Calibration

Assessing the performance of the model for all BSU levels is necessary before comparing scenarios. In *UrbanSim*, as in other LUTI models, this calibration procedure consists in comparing the situation forecasted by the model with the observed reality on a given time step (Wegener and Furst, 1999b; Bonnel et al., 2014). Observed data come here from the 2011 population census (see DGSIE, 2015a). The correspondence between the predicted (from the baseline scenario) population densities in 2011 and this reference is given in Table 6.2.

Indicator	BSU Stat. wards	Sections	Former muni.	Municipalities
(Observed - Predicted inhab/km^2)/(Observed inhab/km^2), in %, in 2011				
10% Quantile	-57.08	-42.66	-6.35	3.02
Median	9.74	8.71	18.23	15.25
90% Quantile	39.27	28.63	36.82	29.93
Observed versus Predicted inhab/km^2, in 2011				
Pearson' ρ	0.91***	0.54***	0.98***	0.99***
Moran' I	-0.001	-0.004	-0.007	0.003

Table 6.2 – **Calibration** (observed data from the 2011 population census; predicted data from the *Baseline* scenario)

Note that the location of the jobs was not available, constraining to limit the calibration to households.

The results in Table 6.2 suggest that the absolute performance of the model is limited, but that tendencies are preserved. The population growth was underestimated during the development of the model, leading to a predicted number of inhabitants of 1 766 947 in 2011 versus 1 906 258 according to the census. Figure E.1 shows that for all BSU levels the model underestimates the future population in the BCR and secondary urban centres, and overestimates it in rural areas. The spatial auto-correlation, however, is not significant (Table 6.2). Overall, the performance appears similar for all BSU levels. This is the critical point here, since it can thus be assumed that eventual variations between BSU levels observed in 2020 will be linked to the scale effect, rather than noises due to a varying goodness-of-fit of the model.

6.4.3 Indicators of location choices, transport, and land-use

Table 6.3 gives the final share of agents within the BCR, for each combination of BSU level and scenario. The standard deviation allows assessing the inter-runs variations. For households, variations are systematically larger between BSU levels than across scenarios, by one order of magnitude. Two groups of BSU levels appear: "small" (*Statistical wards* and *Section*) and "large" (*Former Municipalities* and *Municipalities*). Intra-group differences are limited (but still larger than across scenarios), while a large gap is observed between "small" and "large" BSUs. On the contrary, no difference is observed for tertiary sector' jobs (services), either between BSU levels or across scenarios. For all jobs, the variations are slightly larger. The gap observed between sections and former municipalities is larger than the differences across scenarios. Inside the group of "small" BSU levels, the differences are, however, as limited than across scenarios. The same is true for the "large" BSU levels.

This opposition between the "small" and "large" BSU levels does not hold for transport and land-use indicators at the scale of the study area (Table 6.4). Variations of the commuting distance are very limited for both BSU levels and scenarios. Given this stability, no or few variations were expected for the commuting time. Table 6.4 shows that from *Statistical wards* to *Former municipalities*, values are indeed close to each other. Nevertheless, commuting time for *Municipalities* are 3 to 4 minutes larger than those estimated for other BSU levels, for reasons that will be exposed in section 6.5. Finally, the share of surface occupied by green areas is stable across scenarios. Larger variations are observed between BSU levels, in particular for the Section compared to other BSU levels.

6.4.4 Generalised-costs analysis

Table 6.5 gives the social welfare level of the scenarios. Housing and transport costs are computed from simulation results, while the income level is constant. Estimates of the implementation cost of the scenarios, and of the direct revenues, are extrapolated from the literature. Since the scenarios are fictive (even if inspired from actual projects), the accuracy of the proposed values is limited, and they should be seen as an order of magnitude.

Efficiency and/or equity of cordon congestion taxes have been assessed in numerous theoretical works, but survey of real-world applications are much scarcer. Anas and Lindsey (2011) estimate the implementation costs in London (256 million €, for a toll area of 22 km^2), Stockholm (206 mil. €, for 30 km^2), and Milan (7 mil. €, for 8 km^2). Geographical and technical differences make, however, difficult to deduce generic values from these examples. In London, the perimeter of the tolled area is of 21 km, leading to a cost of 12 million €/km. In Stockholm, it is of 6.8 million €/km (perimeter of 30 km), but the insular nature of the city strongly reduces the number of access points. Annual operation costs (including maintenance and investments) are far from being negligible. Anas and Lindsey (2011) estimate them to 245 (London), 31 (Stockholm), and 15 (Milan) million €. For Stockholm, Eliasson (2009) attributes the lower operating cost (compared to Oslo) to the use of a more automated system (note that a large amount of the operation cost was devoted to call centres and to provision for complaints or legal action). Hence, we have assumed an implementation cost of 10 millions €/km for the BCR which gives a total for the *Cordon* scenario of about 700 million €. In the absence of reliable figure for Belgium, operating costs had been set to 100 million €/year (value derived from those for Stockholm, see Eliasson, 2009). The revenues will be computed from simulations results. Note that since a closed city is assumed,

Indicator	BSU	Scenario *Baseline*	*Cordon*	*Office*	*Subsidy*	*Land-use*
Households	(1)	51.23	51.23	51.26	51.24	51.24
		(0.06)	(0.05)	(0.06)	(0.06)	(0.06)
	(2)	51.07	51.07	51.05	51.08	51.08
		(0.02)	(0.00)	(0.01)	(0.01)	(0.02)
	(3)	54.96	54.98	54.91	54.93	54.93
		(0.08)	(0.08)	(0.06)	(0.07)	(0.07)
	(4)	54.84	54.84	54.85	54.85	54.85
		(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Tertiary jobs	(1)	72.34	72.33	72.35	72.31	72.33
		(0.04)	(0.03)	(0.04)	(0.06)	(0.04)
	(2)	72.34	72.33	72.28	72.30	72.32
		(0.02)	(0.03)	(0.04)	(0.04)	(0.05)
	(3)	72.34	72.32	72.36	72.36	72.35
		(0.05)	(0.03)	(0.07)	(0.04)	(0.05)
	(4)	72.35	72.32	72.32	72.33	72.30
		(0.06)	(0.03)	(0.03)	(0.05)	(0.03)
Total jobs	(1)	69.12	69.13	69.12	69.13	69.11
		(0.03)	(0.01)	(0.02)	(0.02)	(0.01)
	(2)	69.15	69.12	69.11	69.13	69.12
		(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
	(3)	69.49	69.49	69.49	69.50	69.50
		(0.03)	(0.01)	(0.02)	(0.02)	(0.01)
	(4)	69.49)	69.48	69.46	69.50	69.48
		(0.02)	(0.02)	(0.02)	(0.02)	(0.03)

Table 6.3 – Share of agents within the Brussels-Capital Region (average value over the 10 runs in 2020 in %; between brackets: inter-runs standard deviation; 1 = Statistical wards; 2 = Sections; 3 = Former municipalities; 4 = Municipalities)

Indicator	BSU	Baseline	Cordon	Office	Subsidy	Land-use
		Scenario				
Travel time (minute)	(1)	38.35 (0.52)	39.97 (0.50)	37.96 (0.22)	38.20 (0.39)	37.98 (0.19)
	(2)	38.44 (0.20)	39.86 (0.28)	38.43 (0.21)	38.39 (0.24)	38.23 (0.16)
	(3)	38.24 (0.19)	39.29 (0.24)	38.36 (0.29)	38.27 (0.36)	38.50 (0.33)
	(4)	42.84 (0.81)	44.54 (1.13)	41.65 (0.21)	43.73 (0.56)	42.09 (0.34)
Travel distance (km)	(1)	19.24 (0.03)	19.20 (0.03)	19.18 (0.03)	19.21 (0.03)	19.19 (0.02)
	(2)	19.09 (0.03)	19.06 (0.04)	19.11 (0.01)	19.12 (0.01)	19.1 (0.01)
	(3)	19.53 (0.05)	19.50 (0.03)	19.52 (0.02)	19.53 (0.05)	19.52 (0.02)
	(4)	19.20 (0.01)	19.18 (0.02)	19.2 (0.02)	19.23 (0.02)	19.2 (0.01)
Resid. area (%)	(1)	23.99 (0.01)	23.99 (0.02)	23.96 (0.02)	23.99 (0.01)	23.99 (0.01)
	(2)	45.20 (0.05)	45.19 (0.08)	45.20 (0.07)	45.21 (0.07)	45.18 (0.06)
	(3)	40.53 (0.04)	40.54 (0.03)	40.55 (0.05)	40.55 (0.05)	40.55 (0.03)
	(4)	41.92 (0.04)	41.89 (0.02)	41.93 (0.01)	41.92 (0.01)	41.93 (0.04)

Table 6.4 – Transport and land consumption indicators (median value over the 10 runs in 2020; between brackets: inter-run' standard deviation; 1 = Statistical wards; 2 = Sections; 3 = Former municipalities; 4 = Municipalities)

the balance between these revenues and the additional transport costs for the residents will be zero.

Information about construction costs of office in Belgium is extremely partial. Based on two recent projects (EUROPA building, 55 000 sq. meter for 240 million €, see EC, 2015; and the new NATO HQ, 250 000 sq. meter for 1.1 billion €, see NATO, 2014), they can be estimated between 4 000 and 4 300 €/m^2 in the BCR. These two examples are, however, "high-end" buildings, with requirements (prestige, security features) not commonly found in basic office space. The construction cost for flats in the BCR was of 2 670 €/m^2 in 2011 (KF, 2012). Hence, a rough estimation of the *Office* scenario implementation cost would be of 1 to 1.6 billion (assuming a price per square meter ranging from 2 500 to 4 000 €). Data on office space rents are more complete, at least for the BCR (from 165 to 285€/m^2/year in 2012, depending on the location, see CBRE, 2012). The annual revenues taken from the implementation of the *Office* scenario can thus be estimated to 81 millions (for a 100% occupation rate). Assuming no inflation and the maximal estimated cost, the amortisation period would be of 19 years, which seems conceivable. Furthermore, a public ownership is assumed for these new office spaces. If these developments' projects were to take place, it is most likely that they will be organised in a public-private partnership, an option that has not been considered here.

Uncertainties are lower for residential scenarios. In 2010, the cost of the *Subsidy* scenario would have been about 160 millions € (12 276 residential real estate transactions in the BCR, for a total price of 3.1 billion €; data from DGSIE, 2015b). And its total cost over the 2010 to 2020 period can be computed from simulation results. The *Land-use* scenario only requires modifications in the land-use scheme by various public administrations. Its implementation cost can thus be considered as null.

On the short term, direct revenues of both *Subsidy* and *Land-use* scenarios are limited to variations in the stock of local taxes collected. Observed levels (Figure 6.4) vary only slightly within the study area. To simplify, the direct revenues taken from the *Subsidy* and *Land-use* scenarios will thus be considered equal to zero.

6.5 Discussion

6.5.1 Consistency and limitations

Several components of the present chapter can be related to previous findings. First, the sensitivity of parameter estimates on the size of the BSU is a known issue. As indicated in section 6.4.1, it is referred as the *scale effect* in the literature on the MAUP (Fotheringham and Rogerson, 2009; Briant et al., 2010). Variations observed here in parameter estimates are consistent with those found in works dedicated to the MAUP (see, in particular, chapter 4). Secondly, We-

				Scenario		
BSU		*Baseline*	*Cordon*	*Office*	*Subsidy*	*Land-use*
(1)	INC	21.39	21.39	21.39	21.39	21.39
	HC	4.91	4.91	4.91	4.91	4.91
	TC	0.21	2.35	0.19	0.21	0.19
	U	16.27	14.13	16.29	16.27	16.29
	C	0	1.70	15.5 to 16.1	1.48	0
	R	0	2.12	0.80	0	0
	SW	16.27	14.55	15.5 to 16.1	14.79	16.29
	Rank	2	5	3	4	1
(2)	INC	21.39	21.38	21.39	21.39	21.39
	HC	4.89	4.90	4.89	4.89	4.89
	TC	0.21	2.21	0.20	0.20	0.19
	U	16.29	14.27	16.30	16.30	16.31
	C	0	1.70	15.5 to 16.1	1.44	0
	R	0	1.98	0.80	0	0
	SW	16.29	14.55	15.5 to 16.1	14.86	16.31
	Rank	2	5	3	4	1
(3)	INC	19.15	19.16	19.15	19.15	19.16
	HC	3.14	3.14	3.14	3.14	3.14
	TC	0.22	1.96	0.21	0.22	0.21
	U	15.78	14.47	15.80	15.79	15.81
	C	0	1.70	15.5 to 16.1	1.38	0
	R	0	2.11	0.80	0	0
	SW	15.78	14.06	15 to 15.6	14.41	15.81
	Rank	2	5	3	4	1
(4)	INC	19.17	19.17	19.16	19.17	19.16
	HC	2.78	2.78	2.77	2.78	2.77
	TC	0.65	2.56	0.58	0.65	0.60
	U	15.74	13.83	15.60	15.73	15.79
	C	0	1.70	15.5 to 16.1	1.37	0
	R	0	1.88	0.80	0	0
	SW	15.74	14.01	15 to 15.6	14.37	15.79
	Rank	2	5	3	4	1

Table 6.5 – **Generalised-cost analysis** (average values over the 10 runs, in billion €; 1 = Statistical wards; 2 = Sections; 3 = Former municipalities; 4 = Municipalities; INC = income level; HC = housing cost; TC = transport cost; U = utility level; C = implementation cost of the scenario; R = direct revenues; SW = social welfare; Rank = rank of the scenario; note that the inter-run' variations does not affect the ranking)

gener (2011a) explores stochastic variations (i.e. inter-run variations) due to changes in the ratio of the number of choices (i.e. agents who relocate) on the number of alternatives (i.e. BSU). His results show that stochastic variations decrease when this ratio increase. Since the number of agents is fixed here, inter-runs' standard deviation should decrease from small to large BSU levels (reduction of the number of alternatives), which is indeed the case, especially for households (Table 6.3).

Thirdly, the low response of the model system to the scenarios is consistent with previous applications of LUTIs in Europe (e.g. de Palma et al., 2008 for Paris; or the MOEBIUS project for Luxembourg, see Lord and Gerber, 2013). The cordon scenario in Cabrita et al. (2015), who uses the same model system than here, is similar to the *Cordon* scenario implemented here. The observed relocation of agents is negligible in both cases, and the influence on commuting time and distances is also limited. On the contrary, the densification scenario of Cabrita et al. (2015) leads to a large relocation of households towards the BCR (+8.5% compared to the baseline). It is, however, based on the very strong assumption that new residential units are constructed within the BCR at the rate of 2% of the total dwellings stock per annum, for five consecutive years. The changes implemented are, therefore, far larger than for our *Subsidy* and *Land-use* scenarios. Overall, the results appear thus consistent with previous works.

It was impossible to take into account the modal share of car and public transport due to technical difficulties causing *MATsim* to crash when the model choice module is activated[1]. This issue should have a limited influence on the *Baseline*, *Office*, *Subsidy*, and *Land-use* scenarios. For the *Cordon*, the workaround was to apply the congestion fee for the entire day, rather than for the morning peak hour only. To avoid the fee, households must thus either relocate inside the BCR, or select a job located outside of it. Both processes that have a reaction time far longer than a change in travel behaviour (Wegener et al., 1986; Simmonds et al., 2013), therefore constraining the low effects observed for the *Cordon* scenario.

One could also argue that scenarios implementing larger deviations from the initial conditions would have triggered a larger response of the model system (as for the densification scenario of Cabrita et al., 2015). This has not been attempted, since it would have contradicted the purpose of this work, in which variations induced by the scenarios are used as references to assess the sensitivity of a LUTI model to the *scale effect*. A more theoretically sound criticism is that a longer simulation period would have allowed slower urban processes, such as land-use changes (see Wegener et al., 1986; Simmonds et al., 2013) to take place. *UrbanSim*, however, is a path-dependant model: the util-

[1]Multiple trials with various sample size suggest that this is not only a memory issue, but provide no further indications. Moreover, it is unclear if the modal choice option has been used in any operational application of the *UrbanSim/MATsim* coupling plug-in.

ity function of agents, defined by the specifications of the location choice sub models, is fixed over the simulation period. Changes should thus start as soon as the scenarios are implemented, and continue until development constraints (e.g. the maximal number of households per zone) are reached. If no variations are visible after 10 years, it is unlikely that they would occur after twenty.

6.5.2 First and second order sensitivity to scale

The main interest is not to focus on Brussels, but to highlight general results of the sensitivity of LUTI models to scale. Variations observed in the location choices of agents (Table 6.3) appear to be linked with the spatial structure of their perceived utility-level. Figure 6.4 shows that households are less concentrated in the BCR than tertiary sector jobs. A mono centric structure seems thus to have a lower sensitivity to the scale effect than a poly centric one. This result is quite straightforward, since a large centre will always emerge from its neighbourhood, whatever the aggregation level, while small centres may be "diluted" for large BSU levels (see chapter 4). Variations in the distribution of agents can be seen as a first order scale' sensitivity of LUTI models. They are induced by the influence of the MAUP on econometric methods, specifically here the sensitivity of DCM to the size of the areal units that constitute their choice set (see Arauzo-Carod and Antolín-Manjón, 2004; chapter 4).

A second order scale's sensitivity is observed in Table 6.4. These indicators are derived from agents' location choices, but their computations require additional parameters that are themselves sensitive to the BSU level. Let's consider the commuting times: *MATsim* uses as origins and destinations the centroid of the zone. For large BSU levels, a high number of agents are therefore concentrated on the same location. Yet, assuming that all agents within a zone are located on its centroid is usually a too strong hypothesis (see e.g. Goodchild and Gopal, 1989) and *MATsim* thus allows randomly distributing them in a buffer centred on each centroid. The width of this buffer was set here as the radius of a circle whose area is equal to the median area of each BSU level (from *Statistical wards* to *Municipalities*: 320, 790, 1 450, and 2 715 meters). Such correction is needed to avoid local congestion effect that would otherwise arise from the concentration of a large number of agents in a small number of origins or destinations. It appears that it works properly for the three small BSU levels, but that for *Municipalities* the width of the buffer is insufficient.

The total residential area constitutes a second example. This indicator depends on the median area of a residential plot. For detached houses and from *Statistical wards* to *Municipalities*, it is of 716, 982, 1 098, and 1 039 square meters. For *Sections*, the median plot size is far larger than those of *Statistical wards*, which explains the greater values observed in Table 6.4. While for "large" BSU levels, the increase in the median plot size is more than compensated by the relocation observed towards the BCR, where plot size are smaller. These findings raise once again the question of spatial aggregation methods (see e.g. Goodchild and Gopal, 1989).

The social welfare level (Table 6.5) changes through BSU levels due to variations of the inputs (location of agents, real estate prices). The scenarios' ranking are, however, not affected. The generalised-costs approach thus provides consistent results through scale. Reasons are that the scenarios are either implemented at the building level (*Office*) or following the border of the municipalities (*Cordon*, *Subsidy*, *Land-use*). It was a natural choice here, since the municipalities are the only BSU level to have an administrative power. The sensitivity of the results to scale remains, however, an open question for scenarios with a more complex spatial footprint. Nevertheless, the generalised-costs relies on aggregate values at the study area level, and large differences are observed in the balance (revenue minus cost) of the scenarios.

6.5.3 Recommendations for operational applications

The results presented here raise several concerns for operational applications of LUTI models. Let us first note that DCMs are used in almost all state-of-the-art LUTI models (see Wegener, 2004; Simmonds et al., 2013). Therefore, the sensitivity to scale observed for *UrbanSim* is likely to be present in other models, even if variations in magnitude remain an open question. It is also probable that other case studies will show a significant level of sensitivity to scale, linked to the initial spatial distribution of agents. Poly centric patterns (i.e. households here) appear more influenced by the size of the BSU than mono centric one (jobs in services). These findings confirm that a good knowledge of the study area is vital for LUTI model projects (Nguyen-Luong, 2008).

The main finding is that the sensitivity to the size of the BSU varies from one output of the model system to another. It is unclear if primary (i.e. the final location of agents) or secondary (e.g. travel times) indicators should be preferred. LUTI models have been criticised since Lee (1973) for producing only the kind of aggregated results that the former one constitutes. The meaning of the latter may, however, be obfuscated by their computation that requires additional parameters varying with the BSU level. Yet, they highlight an additional dimension of the urban realities. Indicators at the agents' level (as in the generalised-costs approach) would allow taking advantage of the disaggregated nature of micro-simulation model. However, as long as the LUTI model is not based purely on individual observations and agents, the spatial biases will persist. And it is unclear if an evolution towards more disaggregation is a desirable path for LUTI models, due to stochastic variations and longer computation times (see Wegener, 2011a), but also constraints on data availability (see Thomas et al., 2015).

The research question of this chapter was: can the scale effect biases policy evaluation? Our results suggest that it depends on the indicator used. If policy makers require actual predictions (e.g. that the commuting time should decrease by 10%), then the answer is yes. Nevertheless, the direction of the variations between scenarios is preserved (the only significant example being the increase of travel time for the *Cordon*, see Table 6.4) through BSU levels. This is also true for the ranking of the scenarios according to their SW level (Table 6.5). Hence, when the model system is used as a simplified reality, to compare options, then the size of the BSU does not seem to be an issue. This result can be related to the discussion between explanative versus predictive models (Shmueli, 2011). It also depends on the objective of the work.

Unhopefully, to our knowledge, LUTI model results often favour actual predictions (see Badoe and Miller, 2000; Bartholomew, 2007; Handy, 2008). Generalizing policy evaluation's methods based on a consistent, multi-dimensional, economic framework (e.g. Hély and Antoni, 2014; Proost et al., 2015) may reduce the risk of wrongheadedness due to the *scale effect*. They do not, however, solve the fundamental problem that LUTI models themselves (and particularly their econometric components) are sensitive to spatial biases (see chapter 3, chapter 4, and chapter 5), for which no easy or straightforward solution exists. We would like to urge here (as Nguyen-Luong, 2008) that a good knowledge of the study area is vital to select the adequate BSU level, according to data availability constraint and the spatial structure of the city.

6.6 Conclusion

This chapter proposes a sensitivity analysis of a LUTI model' outputs to the size of the areal units used by the model. Using an empirical case study (Brussels, Belgium) and four BSU levels, the results show that variations with scale are generally larger than those between scenarios. A poly centric structure appears more sensitive than a mono centric one, but variations of the results are not monotonous with the BSU level and thus cannot be easily generalised. The main reason is that econometric methods in LUTI models are sensitive to the *scale effect* of the MAUP, confirming previous findings of chapter 4. These results have important implications for policy evaluations based on LUTI models. Actual predictions are likely to be biased by the BSU chosen for the model, especially if separated indicators are used rather than a unified economic framework (cost - benefit analysis). On the contrary, when the model is used as a simplified reality to compare scenarios, the results are consistent through scales. Together with those of chapter 5, these findings call for a better awareness of potential spatial biases in operational applications of LUTI models (see chapter 7). Space may biases LUTI models and, therefore, space matters in their operational applications.

Part IV

Recommendations and conclusion

CHAPTER 7

Recommendations and conclusion

7.1 Executive summary

Different experiments have been conducted in this doctoral dissertation to meet its general objective: assess if the behaviour and outputs of LUTI models are affected by their spatial extent and resolution. The answer is yes. Despite the limitations highlighted in all chapters, our findings demonstrate the sensitivity of LUTI models to space, and provide insights to understanding the causes of this sensitivity, which are both internal (i.e. due to the mechanism of *UrbanSim*) and external (i.e. due to modellers' choices).

The added value of this thesis to the current state-of-the-art is twofold. Part II provides an extension to the existing literature on the MAUP, by assessing the sensitivity of regression methods to the *spatial extent*, and the sensitivity of a discrete choice model to the *spatial resolution*. The analyses conducted in Part III are original (no other published work has considered the question) and demonstrate the influence of spatial bias on the outputs of LUTI models. Overall, a significant sensitivity to spatial bias is found for four of the situations explored in this thesis, and a limited sensitivity for the two others (see Table 7.1). Let us recall that the *behaviour* of LUTI models refers here to their internal principles (i.e. regression methods and DCM), while the *outputs* consist in the final situation predicted by the model. The *spatial extent* designates the size of the study area on which a LUTI model is applied, and the *resolution* the size of its Basic Spatial Units.

Before summarising how they affect LUTI models, let us recall that our main methodological choice (see chapter 1) was to focus on the land use component of LUTI models (represented here by the *UrbanSim* model) and that only two case studies have been used (Brussels and a synthetic city). It is, therefore, necessary to question the validity of our findings for other LUTI models. Chapter 2 shows that other state-of-the-art LUTI models also rely on utility-maximising methods to forecast agents' location choices, and their parameter estimates will be sensitive to spatial bias. Secondly, bias occurring in the data collection and processing steps will be present whatever the modelling framework used. It is thus likely that any LUTI model will show some levels of sensitivity to spatial bias.

7.1.1 Spatial extent

The *spatial extent* has a limited influence on parameter estimates of regression models for real estate prices (see chapter 3). No significant variations are found in parameter estimates. It can, however, be due to the spatially aggregated data used here. The magnitude of the variations of these parameter estimates can be related to changes in the nature of the study area, from the CBD to a mono centric urban region and, eventually, a poly centric metropolitan area. Significant variations of parameter estimates are, on the contrary, observed for Discrete Choice Model (DCM) forecasting agents' location choices (see chapter 4). For both methods, these findings imply that both the size and the composition of the study area influence the socio-economic process occurring within it.

The outputs of *UrbanSim* are influenced by the size and composition of the study area, as shown by the sensitivity analysis on the mono and poly -centric configurations of the synthetic case study (chapter 5). As for econometric components, the composition appears more critical than the size. The inclusion of rural areas, with no or few relationships with the CBD, has only a limited influence. The main risks of bias occur when the study area encompasses a portion of the catchment area of other CBD (i.e. municipalities that are more attracted by a CBD located outside the study area than by those included). The potential influence of the *spatial extent* on policy evaluation has not been assessed in this thesis. Chapter 5 show that simple indicators would have been affected but the sensitivity of a generalised costs approach remains an open question. Our guess is that this potential influence would be linked to the nature of the changes implemented, but the variety of the possible scenarios makes formalisation difficult.

Component	Spatial choices: Spatial extent	Spatial choices: Spatial resolution
Regression analysis	**Limited sensitivity.** See chapter 3: variations of parameter estimates are observed but are not significant. The relevant independent factor vary with the nature of the study area	**Not assessed in this thesis.** Strong evidence of significant sensitivity in the literature; see chapter 1.
Multinomial logit	**Significant sensitivity.** See chapter 5: variations of parameter estimates and utility level.	**Significant sensitivity.** See chapter 4: variations of parameter estimates and utility level linked with the spatial structure of the agents.
Outputs	**Low to significant sensitivity.** See chapter 5: depends on the indicator considered and on the nature of the zone added to the study area.	**Low to significant sensitivity.** See chapters 5 and 6: depends on the indicator considered and on the spatial structure of the agents.
Policy evaluation	**Not assessed in this thesis**	**Limited sensitivity.** See chapter 6: significant variations observed for some simple indicators but no influence on a cost-benefit analysis.

Table 7.1 – Summary of the findings

7.1.2 Spatial resolution

The influence of the *spatial resolution* on the real estate price sub model was not assessed, due to a lack of data. Nevertheless, there is strong evidence from literature (see chapter 1) that parameter estimates of regression methods vary with the size of the BSU. Literature also highlights a similar sensitivity for DCM (see chapter 1). This thesis allowed extending the state-of-the-art in that field by considering smaller BSU than those used in existing works, and by relating directly the sensitivity analysis to LUTI models. Significant variations are found in parameter estimates, affecting in turn the utility level of each BSU. The magnitude of this influence appears to be linked with the spatial structure of agents: a strong mono centric distribution is, indeed, less affected by changes in the size of the BSU than a poly centric one. On a conceptual point of view, this sensitivity to the *spatial resolution* can be related to a varying importance through scales of the process and amenities driving agents' location choices.

The outputs of *UrbanSim* are sensitive to changes in the size of the BSUs (see chapter 5 and chapter 6). Our results show that this sensitivity is a function of the spatial distribution of agents and, therefore, varies from one indicator to another. The explanation of this sensitivity lies in variations of the relative intensity (or even direction) of feedback effects accounted for by the model. Regarding policy evaluations, a limited sensitivity is found. Simple indicators vary with the size of the BSU. Policy evaluation based on threshold values or multi-criteria analysis will, therefore, be biased. Generalised-costs approach, on the contrary, produce consistent results whatever the size of the BSU (i.e. the ranking of the scenario is not affected, even if their final social welfare level changes; at least for the scenarios tested in chapter 6).

7.1.3 Operational implications

Does this sensitivity to spatial bias have implications for operational applications of LUTI models? Again, the answer is yes. Despite limited in magnitude, the variations observed between different spatial extents or resolutions are larger than between scenarios. Moreover, predicting these variations prior to running the simulations is impossible.

The variations of parameter estimates cannot be predicted (see chapters 3 and 4), since the set of independent variables imputed to one econometric sub model is user-defined and, therefore, specific to each case study. Chapter 5 shows that studying the variations of the utility level predicted by the location choice sub models is a more promising study path. Nevertheless, although identifying the portions of the study area where the growth of (for instance) the population will increase or decrease between the different *spatial extent* or *resolution* seems possible, we cannot predict the magnitude of these variations. The reason is that LUTI models account for a large number of different feedback effects, whose intensity and direction are affected by the *boundary* and *scale effects*. Even the very simple specifications used in our synthetic case study lead to unexpected evolutions (see chapter 5). In real-world applications, where agents' location choices depend on multiple variables, the number of feedback effects could be exponentially larger, precluding any disentangling.

Hence, the sensitivity of LUTI models to space is not only a technical problem, but results also from the choices made by the modellers. In the remaining of this chapter, we explore, therefore, three perspectives for future developments. Section 7.2 assumes that we are on the verge of developing operational applications of LUTI models, and attempt to propose “best spatial practices” for the selection of both the *spatial extent* and *resolution*. Section 7.3 then examines the technical development of LUTI models that would improve the representation of space (spatial econometrics methods and multi-scale models). Finally, section 7.4 proposes an alternative approach for land-use and transport modelling based on (1) a better integration of economic geography theory into LUTI models and, (2) the potential development of "feasible LUTI models". Section 7.5 constitutes the general conclusion of the thesis.

7.2 Best spatial practices in LUTI models

7.2.1 How to delineate the study area?

The *spatial extent*, i.e. the size and shape of its study area, affect various dimensions of a LUTI model application (see section 7.1). Especially if the study area encompasses areas under the influence of another city than the main CBD. Hence, the optimal delineation of the study area of an operational application of a LUTI model should present three characteristics. First, interactions between places, through transport, should exist inside the study area. Secondly, the study area should be large enough to assess the influence of large-scale scenarios (e.g. transportation network improvements). Chapter 2 shows quite clearly that LUTI models' applications favour such large study areas, corresponding to the extended urban area category (see Table F.2). Finally, *UrbanSim* (as many other LUTI models) assumes a closed city. To reduce potential bias, interactions between the study area and the “rest of the world” should thus be as limited as possible.

Methods for cities' delineations

This issue relates to the question of the definition of cities' influence area. Due to the long-existing sub urbanisation process, built-up fringes show irregular patterns, recent detached housings estates being mixed with traditional rural buildings and employment opportunities (Tannier et al., 2011). The binary distinction between cities and countryside is thus no longer valid (Schuler et al., 2009). Administrative boundaries are also often obsolete, many cities sprawling out of their official limits (Dujardin et al., 2007). Hence, the urban phenomenon should be studied using morphological or functional delineations taking into account the economic influence of a metropolitan area (Cheshire and Gornostaeve, 2002; Cheshire, 2010).

Official functional delineations of urban areas exist, in various countries (e.g. France, see Le Jeannic and Vidalenc, 1997; Julien, 2000; Julien, 2001, the US, see Federal Register, 2000; Federal Register, 2002; and Switzerland, see Le Gleau et al., 1996; Dujardin et al., 2007). In addition, the methodology followed by Van Hecke et al. (2009) for Belgium is detailed in chapter 3. Most of these functional delineations can be organised in three nested categories (Table F.1), varying clearing in meaning and according to the criteria taken into account. The main drawback of these classical methods for cities' delineation is that they rely on arbitrarily definedthreshold values (Coombes and Bond, 2007; Blondel et al., 2010; and Thomas et al., 2013).

Recognising this potential bias, innovative methodologies have emerged, often relying on interaction matrices among locations. Commuting patterns are particularly favoured, due to their natural relation with labour market areas. Karlsson and Olsson (2006) propose a review of theories and methods for the delineation of such functional regions. Applications based upon the partition of networks representing commuting fluxes can be found for Czech Republic (Klapka et al., 2014), Ireland (Farmer and Fotheringham, 2011), Slovenia (Konjar et al., 2010), United Kingdom (Coombes, 2013), and Brussels (Thomas et al., 2013). Networks of phone calls have also, but less frequently, been used. Notable examples include United Kingdom (Ratti et al., 2010) and Belgium (Thomas et al., 2013). These recent developments in urban delineation' method attempt to identify functional region based upon an endogenous parameter, usually some kind of modularity, i.e. a measure attempting to maximise the intra-region similarity and to minimise the inter-region dissimilarity. Note that delineation of the morphological agglomeration using fractals (see Tannier et al., 2011 for theoretical background, and Tannier and Thomas, 2013 for an application to various European cities) are not discussed here since we are focusing on the complete extent of the labour market area.

Best delineation for LUTI models

Among the indicators commonly used in cities' delineation (Table F.1), we recommend using commuting flows, and to apply a methodology based on the partition of an origin – destination matrix, with an endogenous criterion that attempts to produce clusters with maximal intra-group fluxes and minimal inter-group fluxes. The job basin proposed in Thomas et al. (2013) constitutes a good example. This approach meets the three desirable characteristics expressed for the study area of LUTI models: (a) commuting is the product of the spatial mismatch between residential locations and employment centres. (b) Commuting fluxes produce large delineations, and (c) the proposed methodology will allow creating a study area with strong internal ties but loose links with the rest of the world, reducing potential border effect bias.

Three technical limitations should be addressed. First, origin – destination matrices of commuting fluxes are generally drawn from one areal unit to another (rather than using individual data). Discrepancies can, therefore, still exist at the edge of the delineation, but this bias should be limited if data is available for sufficiently small areal units (practically speaking, no larger than Belgian municipalities). Secondly, it does not guarantee that the fluxes to or from the outside of the study area will be negligible. The need for a better representation of the interactions with the "rest of the world" will, therefore, remain (see section 7.3). Finally, this methodology is only valid for urban applications of LUTI models, i.e. when the study area is centred on one metropolitan area. Regional applications can also be found, i.e. an application of LUTI model encompassing a network of cities or even an entire country (e.g. TIGRIS XL for The Netherlands, see Zondag and de Jong, 2011; Zondag et al., 2015 and MARS for Austria, see Pfaffenbichler et al., 2010). This latter case should be dealt with by relying on a multi-scale model (see section 7.3).

The main limitation is the existence of political constraints. Chapter 2 shows that the *spatial extent* of a LUTI model is not always a matter of choice for the modellers. In this situation, two cases may appear. (a) If the commuting area is larger than the mandatory study area, the poly centric example from chapter 5 shows that the biases should be limited. (b) The opposite situation can, on the other hand, lead to major discrepancies. Yet, this latter case is similar to the regional applications of LUTI models, and should also be dealt with by a multi-scale model.

7.2.2 What zoning system should be used?

Most LUTI models are not purely disaggregated. A zoning system has, therefore, to be defined. The desirable characteristics of such areal units are that (1) they should be consistent with the neighbourhood taken into account by agents in their location choices, but also (2) be useful for policy makers. Zones corresponding to (or that can be re-aggregated into) official administrative boundaries are, therefore, desirable. These two characteristics are somewhat contradictory. Moreover, the *spatial resolution* issue is also technical, and can be divided between constraints external to the model systems, and those inherent to its internal principles.

External constraints

Data availability constitutes the main constraint on the *spatial resolution*, as shown in chapter 2 and section 7.1. Due to the inherent bias of the aggregation or disaggregation procedures (Goodchild and Gopal, 1989; Fotheringham and Rogerson, 2009; Fotheringham et al., 2000a), it is preferable that modellers rely on the areal units for which most of the variables required are available.

A more conceptual issue is that the aim of a LUTI model's application may also influence the optimal level of areal units. Chapter 6 shows that the specifications of the location choice sub models are affected by the size of the areal units. The principal factors (population or jobs' density, accessibility) are selected for all BSU levels, but specifications for large ones generally involve fewer variables (see appendix E.1). In other words, the processes relevant to predicting location choices vary through scales, which is consistent with the findings of Guo and Bhat (2004); de Palma et al. (2007) and Guo and Bhat (2007). Hence, the *spatial resolution* will affect the processes that the model will be able to simulate. If detailed scenarios have to be tested, then small areal units are required. On the contrary, relatively large areal units appear sufficient if only simple forecasts are requested.

Internal constraints

Chapter 2 outlines a link between the internal complexity of LUTI models and their level of spatial disaggregation. Disaggregated LUTI models allow to take advantage of the increasing availability of micro-level data. Nevertheless, smaller areal units mean an increased level of stochastic variations in the outputs of the model (Wegener, 2011a). The most evident solution to these biases is to run the model several times, and to average the outputs of the successive simulations (however increasing the computation time).

A second internal constraint is that the areal units are the same for all categories of agents. Meaning that the neighbourhood considered by households to assess the environmental amenities of a given residential location is identical to the one used by firms to forecast their future locations. It is, however, in contradiction with the desirable features highlighted at the beginning of this section (this issue is discussed in section 7.3).

Overall, the current section mainly raises more questions that modellers should ask themselves prior to the data collection and processing steps. The reason is that the *spatial resolution* depends on the modelling framework used, on data availability, and on the final goal of the work. These components are specific to each case study. Making further generalisations is, therefore, difficult.

7.3 Toward an optimal spatial model

7.3.1 Spatial econometrics: the solution?

The potential of spatial econometrics methods for LUTI models is sometimes discussed (e.g. Löchl and Axhausen, 2010; Efthymiou and Antoniou, 2013; Thomas et al., 2015) but the first operational applications have yet to emerge. Broadly speaking, spatial econometrics methods (see LeSage, 1999; Anselin, 2002; Le Gallo, 2002; LeSage and Pace, 2009; or Anselin, 2013 for a detailed overview) aim at reducing the biases arising from the presence of spatial dependence among dependant or independent variables of a statistical model, or within its error term. Some authors have argued that such methods were mostly useless, since they could prevent one from identifying the true causal parameters (Gibbons and Overman, 2010). Another limitation is that they consist in treating the symptoms (i.e. the spatial dependence problem) rather than the causes (Le Gallo, 2002).

However, the context of LUTI models presents two specificities. First, intrinsic characteristics of agents and residential or non-residential buildings are limited. Both the real estate prices and location choice sub models will thus mostly depend on locations' amenities. Some level of spatial auto-correlation has, therefore, to be expected. The zoning system may induce the same issue, since data availability often constrains to rely on statistical units that do not follow the "natural" boundaries of real estate settlements. Hence, the presence of spatial auto-correlation among econometric components of LUTI models is the product of the internal principles of the model system. Willing to correct that issue by implementing proper statistical methods is, therefore, natural. We briefly present in this section the type of spatial econometric methods that could be implemented within LUTI models, and the advantages and drawback of such technical developments.

Regression methods

UrbanSim (among other LUTI models - see chapter 2) relies on regression analysis to forecast the evolution of real estate prices throughout the simulation period. Numerous theoretical and applied publications exist on Spatial Auto-Regressive model (SAR), and Spatial Error Model (SEM), e.g. LeSage and Pace (2009); Anselin (2013). But also more complex specifications, such as Simultaneous Autoregressive model (Bivand et al., 2013; McMillen, 2003; Elhorst, 2010), or Spatial Durbin Model. The reader interested in further details on these specifications, and on the choice between them, can refer to Le Gallo (2002); Anselin (2002); Anselin et al. (2006); LeSage and Pace (2009); Elhorst (2010); Anselin (2013); or Bivand and Piras (2015).

Many operational applications can be found in the field of hedonic estimations of real estate prices, e.g. Bowen et al. (2001); Wilhelmsson (2002); Anselin and Lozano-Gracia (2009); Löchl and Axhausen (2010); Cavailhès and Thomas (2013); and Pholo Bala et al. (2014). Their implementation within LUTI models would not raise major technical challenges. Both above-mentioned methods are currently available in many statistical analysis packages, including python libraries (PySAL, see Rey and Anselin, 2007) – the language in which the source code of *UrbanSim* is written. The main caveat is that they require a spatial weight matrix. Various specifications exist (see Getis and Aldstadt, 2004; Zhou and Lin, 2008). The literature proposes some methods to select the "best" spatial weight matrix (see Kostov, 2013; Seya et al., 2013) but it would still add a level of complexity to LUTI models. Let us note, however, that LeSage and Pace (2009) show that the variations in parameter estimates from one spatial weight matrix to another are generally limited.

Overall, spatial econometric methods are mature enough for regression analysis. No practical difficulties exist for their implementation within LUTI models. In particular, the selection of the relevant specification (SAR, SEM, or SAC) can be automated by relying on the information provided by the Lagrange multiplier test (see Anselin, 1988a). The only additional work for the modeller will be to select the type of spatial weight matrix, for which default settings can be implemented. Hence, even if their usefulness for solving the *spatial extent* and *resolution* issues remains unknown at this stage, we believe that implementing spatial regression methods within LUTI models is worth the (limited) effort required.

Discrete choice model

The situation is more complex for discrete choice models. Within LUTI models, they rely, for computational tractability purposes, on the classical linear-in-parameter, utility maximisation multinomial logit (MNL) model with random sampling of alternatives (see chapter 4 and McFadden, 1978). Moreover, LUTI models use a zoning system composed of discrete areal units. (Note that revealed preferences' data sets have thus to be used, i.e. data sets with the actual location of the firms, rather than the more classical stated preferences' framework - see Wardman (1988) for a comparison of these approaches). Hence, the typical choice set consists in census tracks or municipalities, and it includes a very large number of alternatives. The presence of spatial autocorrelation among these alternatives is, therefore, a common problem, and the Independence of Irrelevant Alternatives' assumption is unlikely to hold in such case (Sener et al., 2011). This bias can be accounted for by Generalised Spatially Correlated Logit (see Guo and Bhat, 2004; Sener et al., 2011), or by including a spatially weighted average to the utility function of each alternative (Alamá-Sabater et al., 2011). These specifications are, however, far more computationally intensive, and not commonly implemented in econometric software.

Broader concerns must also be mentioned. First, the choice set can be different among agents (Thill, 1992), and especially for residential location choices (see Pagliara and Wilson, 2010 for review). The random sampling of alternatives assumes a perfect knowledge of all alternatives, which is unrealistic given the limited capacity of agents for gathering information (Fotheringham et al., 2000a, Meester and Pellenberg, 2006). Secondly, for small areal units, the relevant extension of the neighbourhood taken into account by agents can exceed the size of these areal units. To correct this bias, a multi-scale modelling structure has been proposed (Guo and Bhat, 2004), but it is itself sensitive to the definition of the neighbourhood (Guo and Bhat, 2007).

Overall, an ideal specification of DCM with spatial choice set would be one that is independent of the level of aggregation in the definition of the zone. That is to say, a model where the probability of a zone i, created by merging two zones j and k, is equal to the sum of the probabilities of j and k, i.e. that $P_i = P_j + P_k$. That equality only holds if the utilities are expressed in logarithm, with $u_i = ln(\beta X_i)$. As for other econometric developments, such specification is, however, far more computationally intensive than the classical linear-in-parameters specification and not commonly implemented in statistical analysis software (Train, 2003).

Therefore, we believe that nested logit models constitute a more feasible approach (see Cornelis et al., 2012 for an application to Belgium). Such framework are implemented in IRPUD and PECAS (see chapter 2) and consist in the selection of one large region, then of the precise location), similar to the urban areas/suburbs/commuting zone/rural areas' typology used by Cornelis et al. (2012). Amenities of a given location should also be computed by buffer of varying bandwidth rather than within the areal unit. This approach does not raise particular technical difficulties, except the definition of the relevant bandwidth.

7.3.2 Multi-scale models

Wegener (2011a) argues that future developments of LUTI models should not consist in increasing the level of detail, but rather to identify the optimal level of detail, on conceptual, temporal, and spatial components. In our opinion, implementing a multi-scale representation of space is indeed more important than increasing the *spatial resolution*, compared to the one of recent operational applications (e.g. Cabrita et al., 2015; de Palma et al., 2015a and Schirmer et al., 2015).

Both IRPUD and DELTA implement a multi-scale representation of space (chapter 2). For IRPUD, it consists in defining large regions surrounding the study area, which is itself divided into small zones (for DELTA, all regions are further divided into zones). The aim is twofold. First, the commuting fluxes between the study area and these external zones are explicitly represented, allowing accounting for the interactions with the rest of the world. Secondly, the population and employment growth can be made endogenous by distributing the total annual increase of households and jobs according to the characteristics of these regions.

The best option to draw the large areal units is, in our opinion, to rely on the same methods of origin - destination matrix partition than the one suggested for the study area itself. The partition method can for instance be run for the whole of Belgium. The cluster encompassing Brussels becomes the study area (and is afterwards divided into small zones), while all other clusters or, at least, that adjacent to the study area, are used as large areal units.

Whether all the large areal units should be further divided into zones (as in DELTA) or only the central one (as in IRPUD) depends on the scope of the work. The former approach is required for regional applications of LUTI models. For instance, the complete extent of our poly centric synthetic city (see chapter 5) could be divided into three large areal units (catchment areas of the two CBDs, and the suburban area). In the relatively common situation (see chapter 2) where modellers are constrained to a regional study area, this multi-scale approach can thus be used to represent the different "life-basin". In an "urban" application of LUTI models similar the mono centric synthetic city and the Brussels case study considered in the thesis, the latter approach is sufficient.

At this point, we do not know the reduction of *boundary* and *scale effects* that a multi-scale approach would allow. It appears, nevertheless, both feasible and theoretically sound. A more careful consideration for geographical structure, compared to current applications, is required in the definition of the multi-scale areal units to fully exploit their potential, but we believe that this framework constitutes a promising path for future developments.

7.4 Another approach?

A remarkable result of recent operational applications of LUTI models (e.g. de Palma et al., 2008; Lord and Gerber, 2013; Cabrita et al., 2015) is the stability of the final situation predicted by the model, whatever the scenarios implemented. The reason appears to be the inertia of the geographic structure of metropolitan areas. Therefore, one can wonder if LUTI models are really useful, given that they produce only limited results but comes at an high cost. Despite these difficulties, including the sensitivity to spatial bias, we believe that LUTI models are likely to remain highly praised tools for policy evaluation in an urban or sustainable development context. For two reasons: (1) the integration between land use and transport is a legal requirement when applying for federal funding for transportation network's improvements in the US (see chapter 2), and (2) thanks to the wide range of indicators that they offer to assess various scenarios (see Efthymiou et al., 2014; Hély and Antoni, 2014; Proost et al., 2015). Hence, it is not useless to ask ourselves on broader questions on the future of LUTI models.

7.4.1 The need for better spatial theories

Theoretical weaknesses of LUTI models

Each component LUTI models can be related to an important field of theory (Figure 7.1) and, when taken independently, is generally recognised as an adequate tool. The challenges arise from the integration of these components within one modelling framework (the LUTI model itself), for three reasons.

First, their econometric components often rely on relatively simple methods compared to the state-of-the-art of the field from which they derive (e.g OLS regression instead of spatial models, see section 7.1; or MNL model instead of nested logit, see Pagliara and Wilson, 2010; Arauzo-Carod et al., 2010 for reviews). We already quoted in chapter 2 that, *"it is strange experience to notice that at symposia on integrated land-use – transport systems often basic principles that were discussed (. . .) [a] considerable time ago are still high on the agenda"* (Timmermans, 2003; pp. 21).

The second reason relates to the urban systems by speed of change (see Table A.2). It suggests that the response time of a land use process to a perturbation is far longer than for transport. Nevertheless, *UrbanSim* assumes a construction time of zero for new real estate developments (see chapter 5). The practice of running the travel model only after a certain interval of land-use model iterations (generally due to computation time constraints) is also counter-factual with the theoretical framework. Finally, if land-use processes are really slow to react, an equilibrium model (adjusting land-use for five or ten year into the future) is sufficient and there is no need for (quasi)-dynamic models. Therefore, a gap appears between the theoretical foundations of LUTI models and the trends observed in their practical implementations and applications.

The third reason is that a link appears between the *spatial resolution* of the model and the economic, demographic, or environmental process that can (or should be) accounted for. The *spatial resolution* affects the processes relevant to predict agents' location choices (see chapters 4 and 5) resulting in variations of the feedback effects implemented in location choices sub models. For instance, the employment location choices sub models of the Brussels case study (see Cabrita et al., 2015) often include, as an independent variable, the population density or the surface occupied by a given activity sector. Chapter 4 suggests that such variables may have counter intuitive influence for small areal units (such as statistical wards) used in the model.

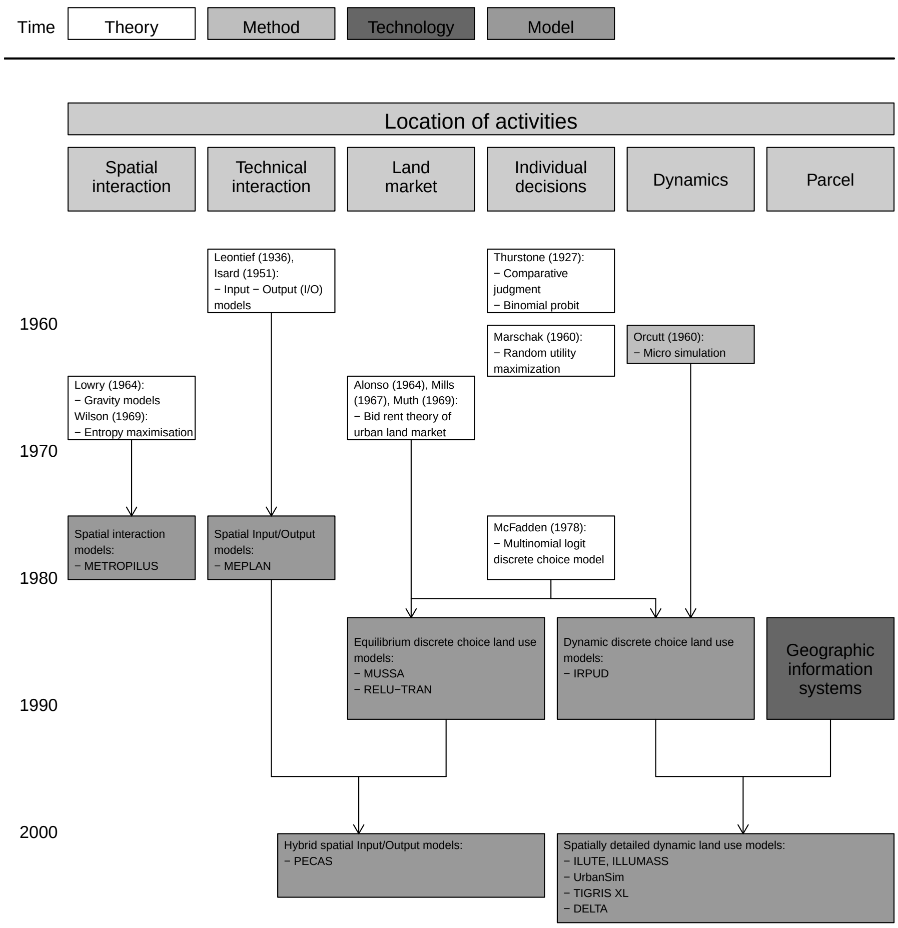

Figure 7.1 – Systematics of land use models (adapted from Zollig Renner et al., 2015)

Overall, LUTI models exhibit a complicated relationship with their theoretical foundations. Various conceptual frameworks are called as justifications for the development towards dynamic, micro-simulation, models. The practices observed are, nevertheless, sometimes contradictory. We believe that there is a need to reinforce the theoretical foundations of modelling choices made in operational applications of LUTI models.

Hidden feedback and consistency

LUTI models attempt to be holistic, i.e. to account for every potential process occurring in metropolitan areas. Nevertheless, they often remain a modular system, in the sense that each sub model is estimated and calibrated independently (Wegener and Furst, 1999b). It causes various methodological issues.

Let us illustrate by an example taken from the Brussels case study of the *SustainCity* project (see Cabrita et al., 2015). The car accessibility is used as independent factor in both the households' location choice sub model (see Table F.3) and the real estate price sub model. The gains in accessibility are thus compensated, in the utility-level perceived by households, by the increase of houses' price that this gain induces. Overall, the specifications of both sub models make sense when considered independently, but result in the inclusion of counter-factual hidden feedback effects in an integrated modelling framework.

On the contrary, urban economics models consist in a set of equations and the feedback effects implemented in the model are, therefore, perfectly explicit. Various applications of micro-economic urban models with a spatially explicit output exist. Former works (e.g. Allen and Sanglier, 1981a; Allen and Sanglier, 1981b; Weidlich and Munz, 1990; and Munz and Weidlich, 1990) focused on the formation of population or employment centres. More interestingly, different recent developments have a scope very similar to the issues tackled by LUTI models:Caruso et al. (2007) and Caruso et al. (2015) have explored the influence of the preference for green spaces leading to suburbanisation. Other examples include the social segregation between rich and poor (Lemoy et al., 2010), the influence of relocation on price formation (Ettema, 2011), the emergence of employment centres (Yang et al., 2012), air pollution and residential location choices (Schindler and Caruso, 2014), or agglomeration economies and transport costs (Delloye et al., 2015).

These spatially explicit urban economics models can provide interesting insight on the feedback effects that should be accounted for in LUTI models. Currently, *UrbanSim* allows the user to define by itself the independent factor of econometric sub models, without guidance. One could imagine an alternative approach consisting in the pre implementation of different feedback effects (e.g. the influence on households' utility level of the real estate prices or of the accessibility to jobs), with either a user-defined intensity (corresponding to the parameter estimate) or fixed relative intensities (e.g. null, low, medium, high). Incompatibilities could also be defined. For instance, going back to the Brussels case study example, the activation of the "accessibility to jobs" factor in the households' location choice sub model could trigger its automatic exclusion from the real estate price sub model. To our opinion, the gain in internal consistency is worth the reduction of the flexibility of LUTI models

Priority	Agents' characteristics Households	Jobs	Buildings characteristics
Mandatory	Number per zone	Number per zone	Prices
High	Income class or level	Local versus non-local activity sector (as in Lowry, 1964)	Maximal number of dwellings and employment floor space per zone
Medium	Size or presence of children	Further refinement into activity sectors	Dwellings' type (houses versus flats)
Low	Age, car's ownership	Home-based status	Dwellings' characteristics

Table 7.2 – Priorities in land use modelling

7.4.2 A faster, more flexible, modelling strategy

Spatial biases excepted, the main message of this thesis is that current state-of-the-art LUTI models are somewhat too complex for practical uses. Hence, we would like to conclude this thesis by a proposal to develop "feasible LUTI models". The research needed to meet this objective can be summarised by the following research question: *"what level of detail is required in the database of a LUTI model to reach a precision level compatible with policy evaluation?"*

Chapters 2 and 6 show that the most cumbersome tasks, in the development of LUTI models' operational application, are the data collection and processing steps (see also Wagner and Wegener, 2007; Nguyen-Luong, 2008). As a result, the simulations are often performed only at the very end of a LUTI model project, reducing the policy evaluation to the bare minimum (see section 6.2 and Wegener, 2011a). Reaching an operational stage earlier during the project would, to our opinion, help modellers in two different ways: (1) allowing to run simple sensitivity analyses to assess the influence of the chosen *spatial extent* and *resolution* on the outputs, and (2) shortening the data processing – simulation – policy evaluation loop, to integrate more closely policy makers and other local stakeholders. This need for a better involvement of these actors is frequently put forward in the literature (e.g. Borning et al., 2008; Hull et al., 2011; Waddell, 2011).

The "feasible LUTI models"' idea combines assumptions grounded in economic geography literature with the automated database generation developed for our synthetic case study (chapter 5). This approach allows generating a "ready-to-use" database in an automated way, as described in section 5.3. It can be applied to real-world metropolitan areas with only limited modifications, by feeding the initial number of households and jobs per zone to the second part of the script described in Table D.3. The usefulness of this proposal should be assessed by three complementary questions.

First, is the loss of precision in the results (compared to a fully disaggregated approach) acceptable? Patterson and Bierlaire (2010) and Patterson et al. (2010) have used aggregated data readily available to build an *UrbanSim* model of Brussels and Lyon. Although their results cannot be easily compared to those of Cabrita et al. (2015), they are *"surprisingly good"* (Patterson et al., 2010; pp. 28). The use of aggregated data *"introduces sufficient noise that only the most robust relationship will manifest themselves in analyses"* (Patterson et al., 2010; pp. 29). Meaning that only the variables grounded in the economic geography literature does have an influence on the agents' location choices. The significant factors for households are the real estate prices, the travel time to the Brussels' CBD, a dummy location for Flanders, and an interaction term of the share of high (or low) income' households with the income level of the household. All these variables are consistent with the Alonso (1964) model. Hence, the work of Patterson and Bierlaire (2010) and Patterson et al. (2010) show that aggregated data can produce a reasonably precise *UrbanSim*' model.

The second question is an application of the Occam's razor principle: which characteristics of agents and land use should be included into the model to reach a sufficiently high goodness-of-fit? Starting from an aggregated model similar to our synthetic city or to the model of Lowry (1964), we attempt to synthesise our answer in Table 7.2. It can be related to Wrigley (1985), who proposes a nested structure of residential location choice model consisting in (a) where to live, then (b) what type of dwelling, and (c) what type of occupant status. One could imagine implementing only the first level in the aggregated version of the case study. The remaining two could be added if the goodness-of-fit is insufficient, or if the required data become available. Note also that the location choices factors emerging from the papers of Patterson and Bierlaire (2010) and Patterson et al. (2010) are consistent with economic geography literature, which is less the case for those implemented by Cabrita et al. (2015), at least for jobs (see Table F.3). Hence providing a further insight that a more disaggregated model may obfuscate the main process driving agents' location choice rather than increasing the goodness-of-fit.

Finally, does a model based on aggregated data, and limited agents' characteristics, produce indicators detailed enough for proper policy evaluation? The social welfare (SW) approach proposed by Proost et al. (2015) relies on four components: net income, housing cost, travel cost of the residents, and global environment. The primary variables used in compute these components, and their availability for the three case studies of the *SustainCity* project, are given in Table F.4. Most of these factors are actually exogenous to the model system. Hence, according to this framework, there is no clear need for more disaggregated or more comprehensive LUTI models than those already existing.

Overall, the "feasible LUTI models" approach appears worth further investigations.

7.5 Concluding words

In this chapter, we proposed recommendations to control or, at least, reduce this influence of the *spatial extent* and *resolution* on the behaviour and outputs of LUTI models. To conclude, we would like to review the relative priority that should be devoted to each of these potential future works.

The first option is to fill the gaps of our experiment plan (see Figure 1.1), i.e. assessing the influence of the *boundary effect* on DCM, and on policy evaluation based on LUTI models' outputs. In a second step, similar analysis to those performed in chapter 6 could be repeated for other LUTI models. DELTA and PECAS appear to be the more relevant candidate, thanks to their various operational applications and their inclusion of some of the desirable feature identified in section 7.2. Such incremental studies are certainly needed. There is, however, little interest in an exhaustive description of the symptoms if no treatment is available.

A more practical approach, therefore, would be to test if the "best spatial practices" proposed in section 7.1. Two main difficulties appear to assess their efficiency. On a practical point of view, it requires for the *spatial extent* to define the optimal extension of the case study, then building an operational application on a study area that exceeds it. Without saying anything of the *spatial resolution* that, as indicated in section 7.1, is difficult to generalise. On a theoretical point of view, we have to define indicators by which this efficiency may be assessed. In other words, performing a calibration procedure far more complete than those conducted in chapters 5 and 6, and probably than those realised in real-world application of LUTI models.

Despite these limitations, the "best spatial practices" are the most spatial of the recommendations detailed in this chapter. It is remarkable that, for a geographer, they are absolutely not groundbreaking. For the rather subjective reason of putting back geography into land use and transport modelling, we would, therefore, recommend assessing in priority their efficiency in any future work devoted on the sensitivity of LUTI models to space.

Section 7.3 highlights various technical developments that could reduce the sensitivity of LUTI models to spatial biases. Their implementation would, however, require much more effort than the adoption of the "best spatial practices". On different occasions, this thesis has pointed out that spatial biases were only one of the various difficulties raised by operational applications of LUTI models. Prior to increasing their internal complexity, one should rather, in our opinion, question the relevance and consistency of LUTI models.

This lead us to the paths for future research proposed in section 7.4. They are essentially speculative. Nevertheless, their key factor is that they do not address spatial biases per se, but rather aim at reducing either the degree of freedom left to the user or the complexity of the model. This thesis has, indeed, confirmed the many flaws of LUTI models already identified in the literature (see Lee, 1973; Lee, 1994; Wagner and Wegener, 2007; and Nguyen-Luong, 2008). Once again, we do not believe that there is a need for more comprehensive (i.e. with a larger number of endogenous process) or more disaggregated models. The main priority in the field of LUTI model should be to develop more intelligent modelling practice, and more theoretically consistent models.

In the epigraph of this thesis, we reproduced a quote from the French film-maker Pierre Schoendoerffer saying that, depending on the type of work produced (novel, film, documentary), an artist is alternatively God, king, or slave. The same is true for modellers. Geographers, because they *"follow and pick-up the traces left behind"*, may be more prone to the latter role than other disciplines. The important point, however, is that one should accept to adapt his role to the context. Modellers can play God when building the internal principles of their model. Nevertheless, as a king is serving its subject, LUTI models should serve their operational applications rather than the opposite. The modelling framework has, therefore, to be the slave of the scope of the study and of data availability. Trespassing these limits is a manifestation of hubris.

Part V

Appendices

APPENDIX A

Appendices of Chapter 2

A.1 The Lowry model

The Lowry (1964) model relies on spatial interactions to forecast the location of future human activities (residential and workplaces). The central assumption is that regional/urban growth is a function of the expansion of a basic sector, that includes all activities meeting non-local demand (practically speaking, all jobs except those in retail). Since the good and services produced by this basic sector are exported outside the study area, it's location and it's evolution over time is assumed to be exogenous and must be given (Rodrigue et al., 2009). Other activities within the study area consist in a retail sector (i.e. the jobs that meet local demands) and a residential sector (i.e. the population). Two nested spatial interaction models (A.1 and A.2) are used to forecast the location of these sectors:

$$T_{ij} = \frac{R_i \times e^{-\beta c_{ij}}}{\sum_i R_i \times e^{-\beta c_{ij}}} \times E_j \tag{A.1}$$

$$S_{ij} = \frac{W_j \times e^{-\beta c_{ij}}}{\sum_i W_j \times e^{-\beta c_{ij}}} \times P_i \tag{A.2}$$

With T_{ij} the work trips from i and j and S_{ij} the shopping trip from i to j. E_j is the number of jobs in j, P_i the number of inhabitants in i, R_i

the number of housing units in i, W_j the shopping facilities in j and c_{ij} the travel cost between i and j (Wegener, 2014). Hence, the core of this model is a spatial-interaction model based on the gravity equation, and the main spatial component is the size (conversely the number) and shape of the BSU among which retail and residential sector are distributed.

A.2 MEPLAN

The MEPLAN model has been developed since the seventies and applied to different case studies (see Echenique, 2001; Hunt and Echenique, 1993). Essentially (Figure A.1), MEPLAN refines the Lowry model by incorporating economic theories and adding a transport sub model with modal choice and assignment (Rodrigue et al., 2009; Echenique, 2001). The Land Use Sub model of MEPLAN predicts the location of employment and residential activities and, as a result, the transport fluxes between the BSU. Employment is divided between two components (exogenous and endogenous), conceptually identical to the basic and retail sector into the Lowry (1964) model. An input - output sub model and an elastic consumption sub model are used to determine the demand for all factors (employment by activity sector and population) in the consumption zones (i.e. the BSU of the model). The spatial allocation process takes this demand as given, and distributes it among all supply sources (Echenique, 2001). The main difference with the Lowry model is that the spatial allocation module uses utility maximization methods to allocate demand to supply (Timmermans, 2003), based on the living cost, the disutility of travel, and the availability of land or floor space in all BSU (Echenique, 2001). This process is summarized by the core equation of the MEPLAN model (A.3):

$$X_{sij} = X_{si} \times A_{si} \times f(c_{si} + g_{sij}) \times Z_{sj} \tag{A.3}$$

With X_{sin} fluxes from region i to region j in the industry sector s, X_{si} the supply of s in i, Z_{si} the demand for s in j, c_{si} the production cost of s in i and g_{sij} the unit transport cost of s between i and j. A_{si} is a factor ensuring that total trade flows from i are equal to the production in i (Wegener, 2014). When the fluxes between BSU have reached equilibrium, the model system calls the Transport Sub model. This step is conceptually the three last stage of a classical four-step model: modal split, route assignment and capacity restraint (Wegener, 2014; see Figure A.1).

A.3 IRPUD

The IRPUD model (see Wegener, 2011b) has been developed for the Dortmund region since 1977. This is a dynamic model, which predicts for each simulation

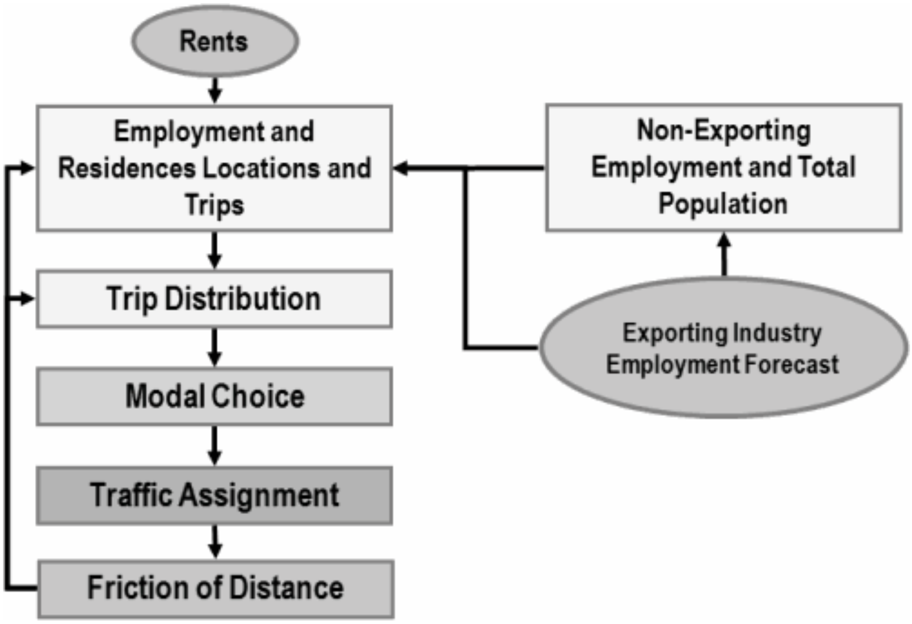

Figure A.1 – The MEPLAN model (figure from Rodrigue et al., 2009)

period (i.e. iteration) location change of households, jobs and real estate developments. For all iterations, IRPUD calls a sequence of sub models (Figure A.2). (1) The transport sub model computes travel time and cost to activities, for different socio-economic groups and by different mode. This transport model is essentially a four-step model with an additional sub model to predict car ownership. The travel demand is determined by a spatial interaction model, as in the Lowry (1964) model but with disaggregation of trips between activities (e.g. to work, for shopping) and socio-economic groups. The other sub models are the following: (2) the ageing sub model estimates the change in the stock of jobs, population and household characteristics for the study area. These evolutions are based on macro-economic trends. Then, (3) the public program sub model is used to implement exogenous events (defined by the user) such as public policy or transport infrastructure investments. These exogenous events are characterized the changes of employment, population or infrastructure that they induce. The (4) private construction sub model simulates the regional land and construction market. Locations of new workplace and dwellings are estimated by random utility maximization. Prices are adjusted at the end of the simulation period, based on observed prices in $t-1$ and on the change in the stock of developable land (by category). (7) The labour market sub model uses a doubly constrained spatial interaction model to forecast job's changes between zones. The decision of changing jobs is a function of the home-to-work trip utility.

Finally, (8) the housing market sub model forecasts the changes of location of the households and the corresponding adjustments in housing prices and rents. A Monte-Carlo micro simulation framework models the housing search process by households. This process is divided into four steps. Multinomial logit models are used to predict if a household looks for a new dwelling (sampling phase) and if yes, in which zones and which types of dwelling he will look for (search phase). The choice of one of the visited dwellings is based upon a threshold in the change of utility. The aggregation phase then multiplies the changes of location by the sampling factor (see Wegener, 2011b for further details). Prices in $t+1$ are adjusted by multiplying the prices in t by the relative change in the stock of vacant dwellings.

A.4 TRANUS

The TRANUS model, developed since 1982, is extensively described in de la Barra (1989). It retains the concept of exogenous production, which corresponds to the basic sector of the Lowry (1964) model, and a spatial input-output model is used to compute the demand of each consumption zone for each sectors. However, as in MEPLAN, the distribution of fluxes to a consumption zone, from production zones, is based on a discrete choice model. The utility (u) of transport, for the sector s, between a consumption zone i and a production zone j is given by:

$$u_{sij} = \lambda_s(p_{sj} + h_{sj}) + t_{sij} \tag{A.4}$$

Where p_{sj} is the price of the sector s in the production region j, h_{sj} the shadow price of the sector s in the region j, t_{sin} the transport disutility (or cost) for goods s between zone i and j, and λ_s a parameter of the relative importance of price compared to transport cost. Note that other characteristics of the zones can be added in the utility function. Another improvement of TRANUS is the representation of prices. At the end of all iterations, an adjustment of prices is performed: if the production within a region i exceeds the maximal production in this region, the prices are increased at the next iteration. Note that TRANUS is an equilibrium model: the convergence in price and production is evaluated at the end of all iterations and the model stops when convergence is reached. As in MEPLAN, the assignation of production to consumption zone is followed by the execution of a transport model to estimate transport cost, modal split and other transport indicators (de la Barra, 1989).

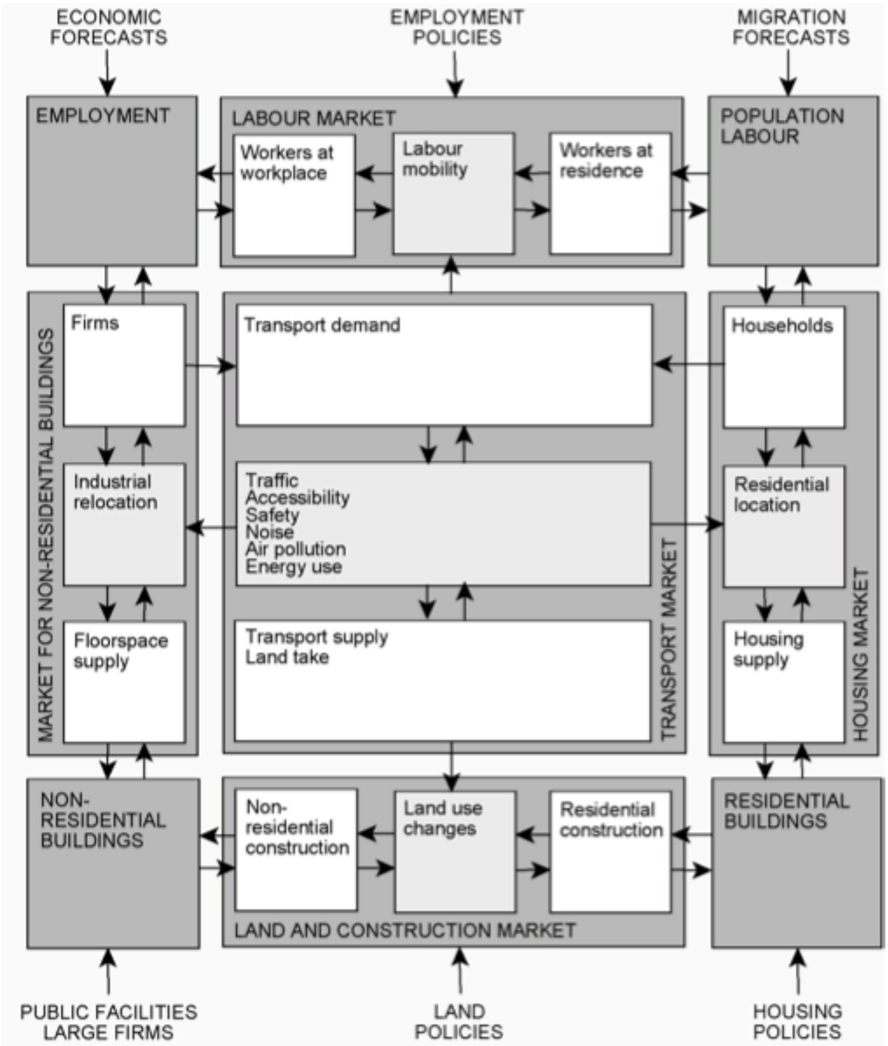

Figure A.2 – The IRPUD model (figure from Wegener, 2011b)

A.5 DELTA

DELTA is a quasi-dynamic modeling package whose development started in 1995. It is based on the START package (see Bates et al., 1991) and is maintained by a consultancy company, David Simmonds Ltd (Timmermans, 2003). A description of the model system can be found in Simmonds and Feldman (2005) and Simmonds et al. (2013).

In the complete version, the DELTA model system calls successively 10 sub models (Figure A.3). Note that DELTA only model land use and have to be interfaced with a travel model that takes land use and computes generalised cost indicators. The sub models are the following: (1) accessibility, which computes the accessibility to different types of work or shopping activities, using logsum (Timmermans, 2003). (2) Development forecasts the location of new real estate. (3) Transition sub model determines the number of new or moving households. (4) Investment performs the same task for employment. (5) Production sub model is a spatial input - output model. (6) Migration predicts long distance change of location among households. (7) Car ownership probability is forecasted by a logit model. (8) Household and Employment location sub models predict the location choices of mobile jobs and households (determined by the transition and investment sub models), using a multinomial logit model. (9) Employment status and commuting sub model use the spatial distribution of households and jobs to predict commuting flows. (10) The housing quality sub model updates the quality of the existing housing stock.

A.6 MUSSA

This model was developed for Santiago de Chile (Martinez and Donoso, 2010). It is described in Martinez (2003) and Martinez and Donoso (2004). MUSSA uses bid-rent and market equilibrium to forecast the location of agents (households, firms). Macro-economic assumptions are used to predict the growth of population and firms over time. MUSSA then forecasts the location of agents for a given point of time in the future, using a static demand-supply equilibrium with location externalities. The model is interfaced with an external travel model that provides accessibility indicators. A multinomial logit model is used to estimates the probability that an agent of cluster h locates in a building of type v in the zone i, conditional to the supply of this building type v in the zone i. The utility is function of a cluster specific constant obtained from the equilibrium solution and of consumer's valuation of the location attributes, either endogenous of exogenous to the model. One particularity of MUSSA is the use of an auction process to distribute the available real estate to the best bidder. The bid represent the consumer's willingness-to-pay and is distributed, within a given cluster of agents, as the utility of a given location plus an identical and independent stochastic term following a Gumbel distribution.

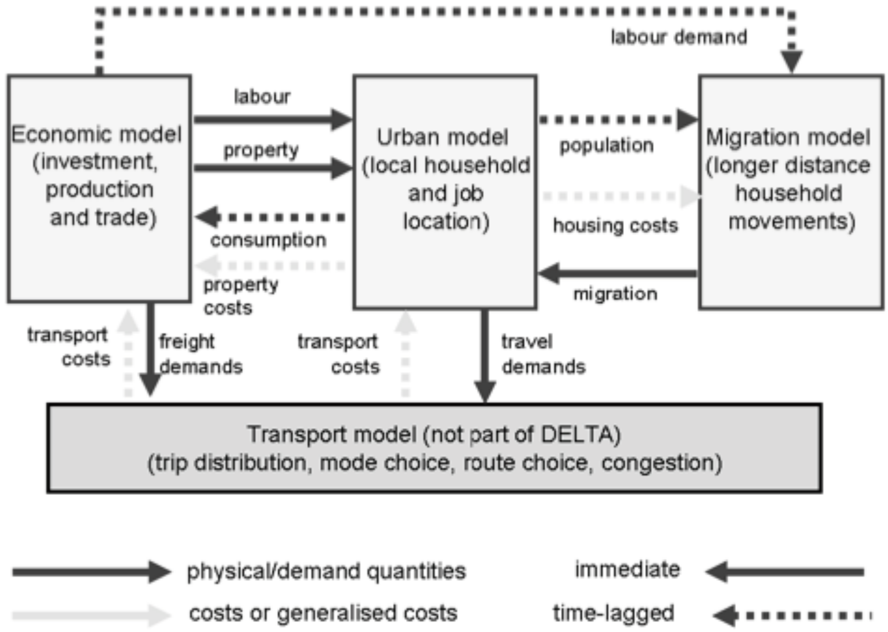

Figure A.3 – The DELTA modelling package (figure from Bosredon et al., 2009)

A.7 PECAS

As MEPLAN and TRANUS, the PECAS model (which stands for Production, Exchange and Consumption Allocation System) relies on a spatial interaction model to determine the location of human activities (Wegener, 2014). Its development started in 2005, making it the most recent of the LUTI models reviewed here. The model system is detailed in Hunt et al. (2005) and Hunt et al. (2009a). Up to date, it has been applied (by chronological order) on Calgary, Oregon, Atlanta, Baltimore, Sacramento, San Diego, California and Alberta . A private consultancy company, HBA SPECTO INCORPORATED, supports the development and applications of the model.

PECAS is a quasi-dynamic model operating though a succession of discrete time steps (typically one year). It combines two internal sub models, Space Development (SD) and Activity Allocation (AA), with two external modules (a transport model, TR, and an economic demographic aggregate forecasting model, ED). Figure A.4 shows the structure of the model. The TR sub model predicts travel time and cost. The choice of the travel model is left to the user (practically speaking, a classical four-step model is sufficient). The ED sub model provides the future number of households and jobs (by category).

This sub model can either consists in static forecast or be able to adjust its prediction to the evolutions simulated by the SD and AA sub models.

Activities (jobs, households) are distributed between Land Use Zones (LUZ) by the AA sub model, using a three-level nested logit model. The first nest allocates a given quantity of each activity to each LUZ. The second nest distributes this quantity between different technological options (i.e. specific production and consumption rates of goods per unit of the activity). The third nest allows to affect's production and consumption to exchange locations. Hence, the share of the activity s allocated to a zone i is a function the location utility of i. In PECAS, it depends on (1) an a priori expected share of i in the total of s, (2) the share of i in s on the previous time-step, (3) an alternative (i.e. LUZ) specific constant, (4) the utility level of each technological options and (5) a set of zonal attributes (i.e. various amenities). Note that in the current implementation of the model, the utility level of these zonal attributes is assumed to be constant over time (see Hunt et al., 2009a; Hunt et al., 2009b, for details).

The SD sub model is available in two versions (aggregated, SD-A, or disaggregated, SD-D). In the SD-A version, each LUZ has a given quantity of land, divided among categories (e.g. high-density residential land, industrial land, etc.), each category being itself divided between developed and vacant space. Space is allocated to category using a multinomial logit function where the utility level is a function of the prices and of the current and available quantity of land. In the disaggregated version (SD-D) of the space development sub model, each LUZ is divided into parcels. A nested logit process is used to select the development events (e.g. no change, new construction, renovation, etc.) affecting each parcel and then the characteristics of this development event (type and quantity of space added or removed). The utility of each development event is a function of the characteristics of the parcel: total area, quantity of developed space, zoning rules defining the allowed category of land use in the parcel, construction cost and prices (or rents).

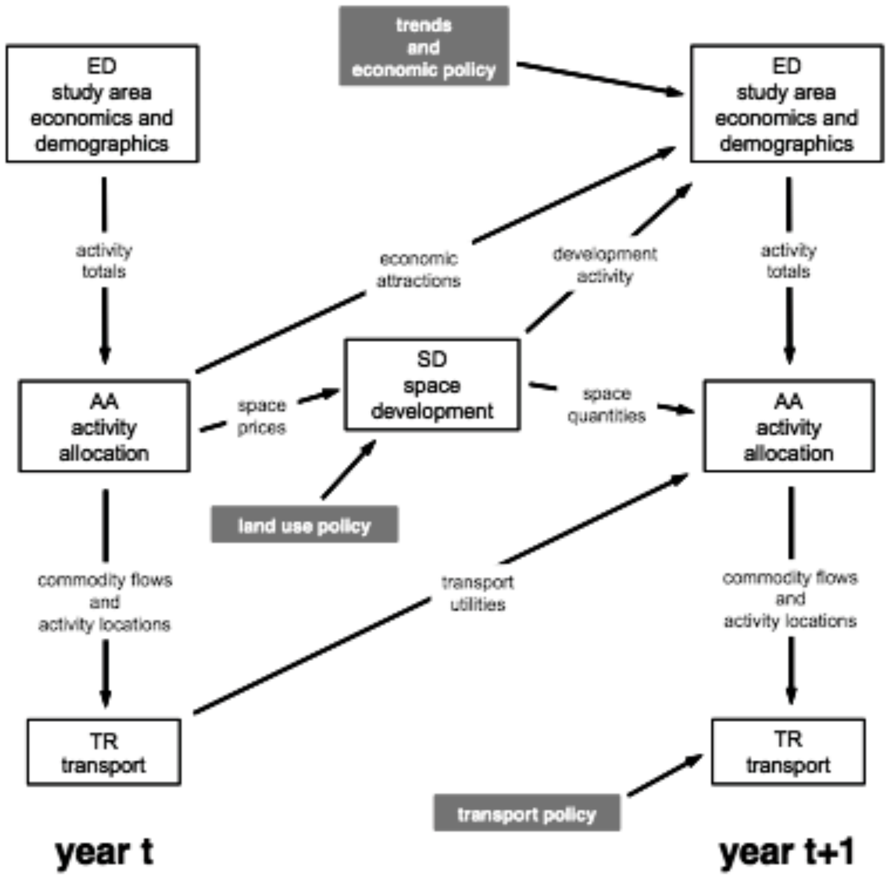

Figure A.4 – The PECAS model (figure from Hunt et al., 2009b)

A.8 Additional tables and figures

Model	Application(s)	Wegener (1994)	Southworth (1995)	Wegener and Furst (1999b)	Wegener (2004)	Hunt et al. (2005)	Simmonds et al. (2013)	Wegener (2014)	Total
MEPLAN	Various (10+)	1	1	1	1	1	1	1	7
TRANUS	Various (10+)	1	1	1	1	1	1	1	7
IRPUD	Dortmund	1	1	1	1	0	1	1	6
MUSSA	Santiago de Chile	1	0	1	1	1	1	1	6
ITLUP	Various (5-10)	1	1	1	1	1	0	0	5
METROSIM	Various (5-10)	1	1	1	1	1	0	0	5
URBANSIM	Various (10+)	0	0	1	1	1	1	1	5
BOYCE	Chicago	1	1	1	1	0	0	0	4
DELTA	Various (10+)	0	0	1	1	0	1	1	4
KIM	Chicago	1	1	1	1	0	0	0	4
LILT	Leeds	1	1	1	1	0	0	0	4
POLIS	San Francisco	1	1	1	1	0	0	0	4
RURBAN	Sapporo	1	0	1	1	0	0	0	3
CUFM	California	1	0	1	1	0	0	0	3
PECAS	South. California	0	0	0	1	0	1	1	3
HUDS	?	1	0	1	0	0	0	0	2
IMREL	Stockholm	0	0	1	1	0	0	0	2
STASA	?	0	0	1	1	0	0	0	2
AMERSFOORT	Amersfoort	0	1	0	0	0	0	0	1
CALUTAS	Tokyo, Nagoya	0	1	0	0	0	0	0	1
HAMILTON	Hamilton	0	1	0	0	0	0	0	1
ILUTE	Toronto	0	0	0	1	0	0	0	1
MARS	Austria	0	0	0	0	0	1	0	1
MASTER	Leeds	0	1	0	0	0	0	0	1
OSAKA	Osaka	0	1	0	0	0	0	0	1
PSCOG	Puget Sound	0	1	0	0	0	0	0	1
TLUMIP	Ohio	0	0	0	1	0	0	0	1
TOPAZ	Australia	0	1	0	0	0	0	0	1
TRANSLOC	Stockholm	0	1	0	0	0	0	0	1
TRESIS	Sidney	0	0	0	1	0	0	0	1
Total		13	17	17	20	6	8	8	30

Table A.1 – LUTI model reviewed over time (1 if the model is reviewed, 0 otherwise)

Speed	Change process	Stock affected	Response time (years)	Response length (years)	Response level	Reversibility
Very slow	Transport construction	Transport networks	5 - 10	$>$ 100	Low	Hardly reversible
	Land use change	Land use pattern	5 - 10	$>$ 100	Low	Hardly reversible
Slow	Industrial construction	Industrial buildings	3 - 5	50 - 100	Low	Very low
	Residential construction	Residential buildings	2 - 3	60 - 80	Low	Low
Medium speed	Economic change	Employment and firms	2 - 5	10 - 20	Medium	Reversible
	Demographic change	Population and households	0 - 70	0 - 70	Low/High	Partly reversible
Fast	Firm relocation	Workplace occupancy	$<$ 1	5 - 10	High	Reversible
	Residential mobility	Housing occupancy	$<$ 1	5 - 10	High	Reversible
Very fast	Change in demand	Goods transport	$<$ 1	$<$ 5	High	Reversible
	Change in mobility	Person travel	$<$ 1	$<$ 1	High	Reversible

Table A.2 – The urban sub systems by speed of change (adapted from Simmonds et al., 2013)

Model	Sub systems modelled	Model theory	Policies modelled
POLIS	Employment, Population, Housing, Land use, Travel	Random utility, Locational surplus	Land-use regulations, Transportation improvements
CUFM	Population, Land use	Location rule	Land-use regulations, Environmental policies, Public facilities, Transportation improvements
BOYCE	Employment, Population, Networks, Travel	Random utility, General equilibrium	Transportation improvements
KIM	Employment, Population, Networks, Goods transport, Travel	Random utility, Bid rents, General equilibrium, Input-Output	Transport improvements
METROSIM	All subsystems except goods transport	Random utility, Bid rent, General equilibrium	Transportation improvements, Travel cost changes
ITLUP	Employment, Population, Land use, Networks, Travel	Random utility, Network equilibrium	Land use regulations, Transportation improvements
HUDS	Employment, Population, Housing	Bid-rent	Housing programs
TRANUS	All subsystems	Random utility, bid rent, Network equilibrium, Land use equilibrium	Land-use regulations, Transportation improvements, Transport cost changes
5-LUT	Population, Housing, Network	Random utility, Bid rent, General equilibrium	Transportation improvements
LILT	All subsystems except goods transport	Random utility, Network equilibrium, Land use equilibrium	Land-use regulations, Transportation improvements, Transport cost changes
MEPLAN	All subsystems	Random utility, Network equilibrium, Land use equilibrium	Land-use regulations, Transportation improvements, Transport cost changes
IRPUD	All subsystems except goods transport	Random utility, Network equilibrium, Land use equilibrium	Land-use regulations, Housing programs, Transportation improvements, Travel cost changes
RURBAN	Employment, Population, Housing, Land use	Random utility, Bid rent, General equilibrium	Land-use regulations, Transportation improvements

Table A.3 – LUTI models' internal principles (summary adapted from Wegener, 1994)

Case study	State	Reference	Delineation	Composition	Surface	Population	Results	BSU
Flagstaff	Arizona	Country (2003)	MPA	City/Township	1 344	135 000	Study Area	Unknown
Maricopa	Arizona	MAG (2002)	MPA	County (1)	25 000	3 800 000	Unknown	Unknown
Phoenix	Arizona	CoP (1997)	Network	Development		0	N/A	Plots
Central Valley	California	AFT (1995)	Admin.	Counties (11)	57 500	4 000 000	Counties	750 000 pixels
Contra Costa	California	CCC (2003)	Admin.	County (1)	2 058	1 000 000	TAZ	TAZ, Pixels
North Livermore	California	ACP (2000)	Network	Development	19	0	N/A	Plots
Sacramento	California	Appendix (?)	MPA	Counties (>1)	17 000	2 500 000	Unknown	Census block, Pixels
San Diego	California	SDAG (1995a)	MPA	Counties (>1)	10 908	2 800 000	Counties	Unknown
San Diego	California	Appendix (1995)	MPA	Counties (>1)	10 908	2 800 000	N/A	N/A
San Diego	California	Appendix (1995)	MPA	Counties (>1)	10 908	2 800 000	N/A	Pixels
San Diego	California	SDAG (1995b)	MPA	Counties (>1)	10 908	2 800 000	N/A	Unknown
San Francisco	California	SDAG (1996)	Admin.	Counties (9)	18 000	7 400 000	Counties	Unknown
Southern Calif.	California	Appendix (2004)	MPA	Counties (6)	98 000	18 000 000	N/A	N/A
Denver	Colorado	DRCG (1995)	MPA	Counties (9)	13 171	2 400 000	Study Area	Unknown
Denver	Colorado	CCoD (2002)	Admin.	County (1)	400	554 000	Neigbourhood	Mod. Census block)
Wilmington	Delaware	WAPC (2000)	Admin.	Counties (2)	2 300	560 000		Census block
Wilmington	Delaware	Appendix (2003)	Development	Development		0	Study Area	Plots
Wilmington	Delaware	WAPC (2003)	Admin.	Counties (2)	2 300	560 000		Census block
Washington DC	DoC	Appendix (1993)	MPA	Counties (7)	14 500	5 000 000	N/A	N/A
Washington DC	DoC	CBF (1996)	MPA	Counties (7)	14 500	5 000 000	N/A	N/A
Washington DC	DoC	CBF (1996)	Development	Development	0		N/A	N/A
Martin & St. Lu.	Florida	TCRPC (2002)	Network	Corridor	500	190 500	Study Area	Census block, Pixels
Orlando	Florida	METRO (2004)	Network	Corridor			N/A	N/A
Atlanta	Georgia	EPA (2000)	Network	Development	0	0	Study Area	Census block, Plots
Atlanta	Georgia	GRTA (2003)	Network	Corridor	2 800		Study Area	Census block
Post Falls	Idaho	CPF (2002)	Network	Corridor	30		Study Area	Plots
Treasure Valley	Idaho	TVFP (2002)	MPA	Counties (>1)	4 000	460 000	Census block	Census block, Pixels
Kishwaukee	Illinois	CRI (2002)	Watershed	Other	135		Plots	Plots (GIS overlay)
Lake county	Illinois	Appendix (1999)	Admin.	County (1)			N/A	N/A
Baltimore	Maryland	Appendix (1992)	MPA	Counties (>1)	6 721	2 500 000	N/A	N/A
Baltimore	Maryland	Appendix (1996)		Unknown			N/A	N/A
Baltimore	Maryland	BRTB (2003)	MPA	Counties (>1)	6 721	2 500 000	TAZ, Counties	TAZ
Chesapeake Bay	Maryland	Appendix (2003)	Watershed	Counties (>1)	165 000	1 7000 000	N/A	County
Montgomery	Maryland	MCPB (2002)	Development	Development		0	N/A	N/A
Detroit	Michigan	Appendix (1993)	Admin.	Unknown			Study Area	N/A
Lansing	Michigan	TCRPC (2003)	Admin.	Counties (3)	4 440	450 000	N/A	Census block
Eureka	Minnesota	Appendix (2003)	Admin.	City/Township	90	1 500	Study Area	Plots
Twin Cities	Minnesota	CEE (1999)	Admin.	Counties (13)	30 000	3 200 000	N/A	Custom
Twin Cities	Minnesota	Appendix (2002)	MPA	Counties (7)	21 000	2 800 000	S.A. + Plots	GIS overlay
Missouri	Missouri	MARC (2001)	Development	Development	2		N/A	Plot

Continued on next page ...

... continuated from previous page

Case study	State	Reference	Delineation	Composition	Surface	Population	Results	BSU
Delaware Valley	New Jersey	DVRPC (2002)	MPA	Counties (9)	9 883	5 500 000	N/A	N/A
New Jersey	New Jersey	NJOSP (1992)	Admin.	State	22 591	8 400 000	Study Area	Unknown
New Jersey	New Jersey	NJOSP (2000)	Admin.	State	22 591	8 400 000	Study Area	Unknown
Princeton	New Jersey	Appendix (1991)	Admin.	Counties (3)	2 219	1 500 000	N/A	N/A
Albuquerque	New Mexico	CoA (2000)	Functional	Other		502 095	Custom (14)	Custom
Albany	New York	Appenix (?)	Network	Corridor			N/A	N/A
Raleigh	North Carolina	GT (1998)	MPA	Unknown			Study Area	N/A
Fargo-Moorhead	North Dakota	FMCoG (1998)	MPA	Counties (2)	7 500	175 000	N/A	N/A
Columbus	Ohio	MORPC (2004)	MPA	Counties (7)	9 343	1 800 000	Study Area	Pixels
Albany	Oregon	Appendix (2001)	Admin.	City/Township	45	50 000	N/A	N/A
Marion County	Oregon	Appendix (2002)	Admin.	County (1)	3 092	315 000	Study Area	N/A
Portland	Oregon	Appendix (1997)	Network	Unknown			N/A	GIS Overlay
Rogue Valley	Oregon	RVCoG (1999)		Unknown			Unknown	Unknown
Salem	Oregon	Appendix (2000)	Admin.	City/Township	125	150 000	N/A	N/A
Willamette Basin	Oregon	ODT (2001)	Watershed	Counties (>1)	30 000	1 100 000	Study Area	122 TAZ
Willamette Basin	Oregon	Annex (2001)	Watershed	Counties (>1)	30 000	1 100 000	Study Area	GIS overlay
Willamette Basin	Oregon	Shermann (1994)	Watershed	Counties (>1)	30 000	1100000	Study Area	Pixels (30 m)
Delaware Valley	Pennsylvania	DVRPC (2003)	Admin.	Counties (9)	9 883	5 500 000	Counties	Unknown
Catawba	South Carolina	CRCG (2003)	Admin.	Counties (4)	6 094	380 000	Counties	TAZ
Pee Dee Region	South Carolina	PDRCC (2003)	Admin.	Counties (6)	9 222	340 000	Counties	398 TAZ
Santee-Lynches	South Carolina	SLRCG (2003)	Admin.	Counties (4)	6 200	2 200 000	Counties	Census block
Oak Ridge	Tennesse	ORNL (2002)	Development	Unknown	20		Study Area	Plot
Central Texas	Texas	ECT (2003)	Admin.	Counties (5)	11 000	1 250 000	Counties	Unknown
San Antonio	Texas	Appendix (2002)	Network	Corridor			N/A	N/A
Mountain View	Utah	CUF (2004)	Network	Corridor			Study Area	Unknown
Salt Lake City	Utah	CUF (1999)	MPA	Counties (10)	60 000	1 600 000	N/A	GIS overlay
Chittenden County	Vermont	CCMPO (2004)	Admin.	County (1)	1 603	150 000	Study Area	TAZ
Hampton Roads	Virginia	HRPDC (2003)	Admin.	Counties (>1)	7 472	1 500 000	Counties	1000+ TAZ
Jefferson	Virginia	TJPDC (2002)	Admin.	Counties (3)	5 550	230 000	N/A	Unknown
Puget Sound	Washington	Appendix (1990)	Admin.	Unknown	16 000	3 200 000	N/A	N/A
Sheboygan	Wisconsin	Appendix (1996)	MPA	Unknown			N/A	N/A

Table A.4 – LUTI projects in the US (sample from Bartholomew, 2007)

City	Authors	Journal	Published in	Mention of size?	Population (millions)	Area (km^2)	Map
Bilbao	Echenique et al.	Trsp Rev	1990	N	n/a (0.95).	n/a	N
Bilbao	Burgos	EP-b	1994	Y	1	n/a	N
Brussels	Patterson et al.	JTLU	2010	Y	2.9	4 361	Y
Brussels	Patterson and Bierlaire	EP-a	2010	Y	2.9	4 361	Y
Dortmund	Echenique et al.	Trsp Rev	1990	N	n/a (0.58)	n/a	N
Dortmund	Mackett	Trsp Rev	1990	Y	1.1	833	N
Dortmund	Wegener et al.	Trsp Rev	1991	Y	2.3	n/a	Y
Dortmund	Wagner and Wegener	disP	2007	Y	2.6	n/a	Y
Edinburgh	May et al.	TRR	2005	Y	2.7	2 305	N
Lausanne	Patterson and Bierlaire	EP-a	2010	Y	0.27	200	Y
Leeds	Echenique et al.	Trsp Rev	1990	N	n/a (0.72)	n/a	N
Leeds	Mackett	Trsp Rev	1990	Y	0.5	164	N
Leeds	May et al.	TRR	2005	Y	2.1	559	Y
Londres	Batty et al.	EP-b	2013	Y	14	n/a	Y
Lyon	Patterson et al.	JTLU	2010	Y	1.6	3 325	Y
Madrid	Guzman et al.	CSTP	2014	Y	6.5	8 000	Y
Madrid	Wang et al.	CEUS	2014	Y	6.5	8 030	Y
Naples	Hunt	EP-b	1994	N	n/a (4.4)	n/a	Y
NL	Eradus et al.	JTG	2002	N	6	n/a	Y
NL	Zondag et al.	CEUS	2014	N	n/a	n/a	N
Oslo	Vold	TR-a	2005	Y	0,95	n/a	N
Paris	Anas	EP-b	2013	N	n/a (12)	n/a (762)	Y
Rome	Di Zio et al.	JTLU	2010	Y	2.6	1 500	Y
Santander	Coppola et al.	JUPD	2013	Y	0.28	n/a	Y
Stockholm	Anderstig and Mattsson	Transportation	1992	N	n/a (2.1)	n/a	N

Table A.5 – LUTI models applications in Europe (based on papers published in peer-reviewed journals)

APPENDIX B

Appendices of Chapter 3

B.1 Additional tables and figures

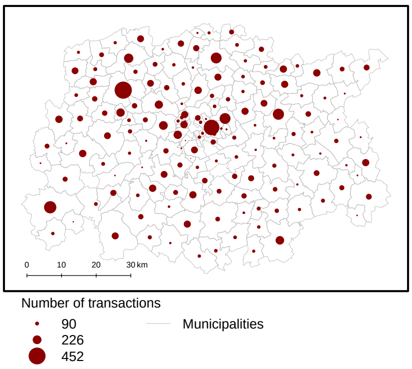

Figure B.1 – Number of real estate transactions (developable land plot, 2006 - 2008)

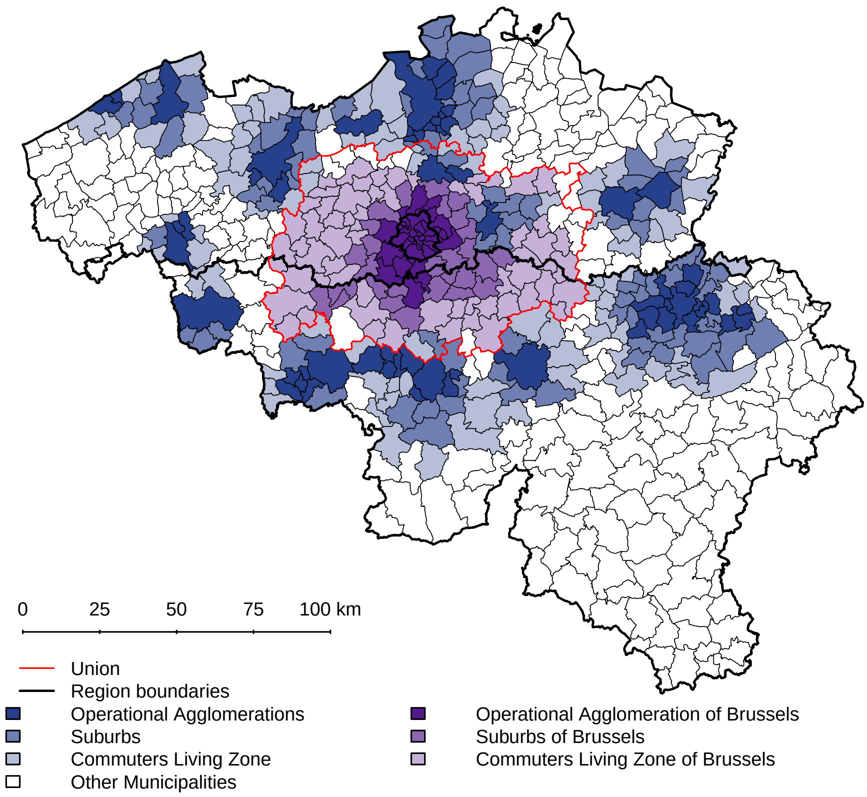

Figure B.2 – Urban region in Belgium (based on the 2001 population census; definition from Van Hecke et al., 2009)

		Control variables				Accessibility indicators	
Delineations	Pop. density	Income	Forest	IC/IR	Wallonia	Travel time to BXL	Accessibility to jobs
BCR	ABC	ABC		ABC		AB	C
OA	ABC	ABC	AC		B		AC
InnRER	ABC	ABC	AB		B		AC
Fractal	ABC	ABC	AB		B		AC
UR	ABC	ABC			ABC		
Phone	ABC	ABC	A	ABC	ABC		AC
Job	ABC	ABC	ABC	ABC	ABC	B	AC
Brabant	ABC	ABC		ABC	ABC	B	AC
MLA	ABC	ABC	B	ABC	ABC	B	AC
RER	ABC	ABC	ABC	ABC	ABC	B	AC
ExtRER	ABC	ABC	ABC	ABC	ABC	B	AC
Union	ABC	ABC	ABC	ABC	ABC	B	AC

Table B.1 – Results of the stepwise regression (A = variable included in the land price model for case A, etc)

		Variables									
ZE	**Case**	Constant	Pop/km^2	Income	$T_{IC/IR}$	% Forest	Wal.	T_{BXL}	A_i	**R^2**	**ME**
BCR	A	-1.31	0.53*	0.21*	-0.15			-0.03		0.51	0.26
(n = 19)		(2.68)	(0.21)	(0.07)	(0.10)			(0.02)			
	B	-1.31	0.53*	0.21*	-0.15			-0.03		0.51	0.26
		(2.68)	(0.21)	(0.07)	(0.10)			(0.02)			
	C	-10.67	0.41	0.17**	-0.2				0.99	0.51	0.26
		(8.65	(0.20)	(0.05)	(0.09)				(0.74)		
OA	A	-9.37	0.52***	0.06*		0.87			0.9	0.72	0.29
(n = 36)		(5.30)	(0.09)	(0.02)		(0.56)			(0.50)		
	B	0.64	0.54***	0.05			-0.37			0.70	0.31
		(1.09)	(0.08)	(0.02)			(0.25)				
	C	-9.37	0.52***	0.06*		0.87			0.9	0.72	0.29
		(5.30)	(0.09)	(0.02)		(0.56)			(0.50)		
InnRER	A	-10.85*	0.5***	0.06**		0.78			1.04*	0.76	0.28
(n = 41)		(4.71)	(0.08)	(0.02)		(0.43)			(0.45)		
	B	0.36	0.56***	0.05*		0.62	-0.39*			0.75	0.28
		(0.97)	(0.07)	(0.02)		(0.43)	(0.19)				
	C	-10.85*	0.5***	0.06**		0.78			1.04*	0.76	0.28
		(4.71)	(0.08)	(0.02)		(0.43)			(0.45)		
Fractal	A	-8.43	0.48***	0.06*		0.7	-0.37*		0.84	0.81	0.29
(n = 48)		(5.24)	(0.08)	(0.02)		(0.43)	(0.17)		(0.49)		
	B	0.37	0.55***	0.05*		0.62	-0.52***			0.80	0.30
		(0.96)	(0.07)	(0.02)		(0.44)	(0.14)				

Continued on next page ...

... continuated from previous page

ZE	Case	Constant	Pop/km^2	Income	$T_{IC/IR}$	% Forest	Wal.	T_{BXL}	A_i	R^2	ME
					Variables						
	C	-8.43	0.48***	0.06*		0.70	-0.37*		0.84	0.81	0.29
		(5.24)	(0.08)	(0.02)		(0.43)	(0.17)		(0.49)		
UR	A	1.79*	0.43***	0.03			-0.57***			0.84	0.29
(n = 62)		(0.68)	(à.05)	(0.02)			(0.11)				
	B	1.79*	0.43***	0.03			-0.57***			0.84	0.29
		(0.68)	(0.05)	(0.02)			(0.11)				
	C	1.79*	0.43***	0.03			-0.57***			0.84	0.29
		(0.68)	(0.05)	(0.02)			(0.11)				
Phone	A	-6.15	0.43***	0.04*	0.03*	0.51	-0.49***		0.70	0.89	0.27
(n = 66)		(4.26)	(0.07)	(0.02)	(0.01)	(0.37)	(0.13)		(0.41)		
	B	1.2	0.48***	0.04	0.02		-0.62***			0.89	0.28
		(0.72)	(0.05)	(0.02)	(0.01)		(0.11)				
	C	-6.15	0.43***	0.04*	0.03*	0.51	-0.49***		0.70	0.89	0.27
		(4.26)	(0.07)	(0.02)	(0.01)	(0.37)	(0.13)		(0.41)		
Commute	A	-8.65***	0.4***	0.05**	0.03**	0.45	-0.45***		0.95***	0.89	0.25
(n = 105)		(2.33)	(0.05)	(0.02)	(0.01)	(0.32)	(0.08)		(0,24)		
	B	1.16	0.47***	0.04**	0.03***	0.53	-0.44***	-0.01**		0.89	0.26
		(0.63)	(0.05)	(0.02)	(0.01)	(à.33)	(0.08)	(0.00)			
	C	-8.65***	0.4***	0.05**	0.03**	0.45	-0.45***		0.95***	0.89	0.25
		(2.33)	(0.05)	(0.02)	(0.01)	(0.32)	(0.08)		(0.24)		
Brabant	A	-7.42***	0.36***	0.03*	0.02**		-0.49***		0.89***	0.88	0.24
(n = 111)		(2.14)	(0.05)	(0.01)	(0.01)		(0.07)		(0.22)		

Continued on next page ...

... continuated from previous page

ZE	**Case**	Constant	Pop/km^2	Income	$T_{IC/IR}$	% Forest	Wal.	T_{BXL}	A_i	**R^2**	**ME**
					Variables						
	B	2.02***	0.41***	0.03*	0.02*		-0.53***	-0.01**		0.88	0.25
		(0.58)	(0.04)	(0.01)	(0.01)		(0.07)	(0.00)			
	C	-7.42***	0.36***	0.03*	0.02**		-0.49***		0.89***	0.88	0.24
		(2.14)	(0.05)	(0.01)	(0.01)		(0.07)		(0.22)		
MLA	A	-9.61***	0.37***	0.04**	0.03**		-0.42***		1.06***	0.91	0.24
(n = 122)		(1.93)	(0.05)	(0.01)	(0.01)		(0.07)		(0.20)		
	B	1.32*	0.46***	0.04**	0.03***	0.43	-0.45***	-0.01***		0.9	0.25
		(0.55)	(0.04)	(0.01)	(0.01)	(0.30)	(0.07)	(0.00)			
	C	-9.61***	0.37***	0.04**	0.03**		-0.42***		1.06***	0.91	0.24
		(1.93)	(0.05)	(0.01)	(0.01)		(à.07)		(0.20)		0.00
RER	A	-8.57***	0.35***	0.03*	0.03**	0.39	-0.42***		1***	0.88	0.23
(n = 126)		(1.86)	(0.04)	(0.01)	(0.01)	(0.26)	(0.06)		(0.19)		
	B	1.9***	0.41***	0.03*	0.03**	0.48	-0.46***	-0.01***		0.87	0.24
		(0.54)	(0.04)	(0.01)	(0.01)	(0.26)	(0.06)	(0.00)			
	C	-8.57***	0.35***	0.03*	0.03**	0.39	-0.42***		1***	0.88	0.23
		(1.86)	(0.04)	(0.01)	(0.01)	(0.26)	(0.06)		(0.19)		
ExtRER	A	-8.06***	0.38***	0.04**	0.03***	0.45	-0.46***		0.92***	0.88	0.23
(n = 147)		(1.75)	(0.04)	(0.01)	(0.01)	(0.25)	(0.06)		(0.18)		
	B	1.44**	0.45***	0.04**	0.03***	0.53*	-0.51***	-0.01***		0.87	0.24
		(0.52)	(0.04)	(0.01)	(0.01)	(0.26)	(0.06)	(0.00)			
	C	-8.06***	0.38***	0.04**	0.03***	0.45	-0.46***		0.92***	0.88	0.24

Continued on next page ...

... continuated from previous page

		Variables									
ZE	**Case**	Constant	Pop/km^2	Income	$T_{IC/IR}$	% Forest	Wal.	T_{BXL}	A_i	**R^2**	**ME**
		(1.75)	(0.04)	(0.01)	(0.01)	(0.25)	(0.06)		(0.18)		
Union	A	-8.33***	0.37***	0.04**	0.03***	0.38	-0.47***		0.95***	0.89	0.23
(n = 160)		(1.67)	(0.04)	(0.01)	(0.01)	(0.24)	(0.06)		(0.18)		
	B	1.38**	0.45***	0.04**	0.03***	0.46	-0.52***	-0.01***		0.88	0.24
		(0.50)	(0.04)	(0.01)	(0.01)	(0.25)	(0.06)	(0.00)			
	C	-8.33***	0.37***	0.04**	0.03***	0.38	-0.47***		0.95***	0.89	0.23
		(1.67)	(0.04)	(0.01)	(0.01)	(0.24)	(0.06)		(0.18)		

Table B.2 – Parameter estimates of the OLS model (significance level: * = $\rho \leq 0,1$; ** = $\rho \leq 0,05$; *** = $\rho \leq 0,01$; between brackets: standard deviation; ME = Median error)

APPENDIX C

Appendices of Chapter 4

C.1 Additional tables and Figures

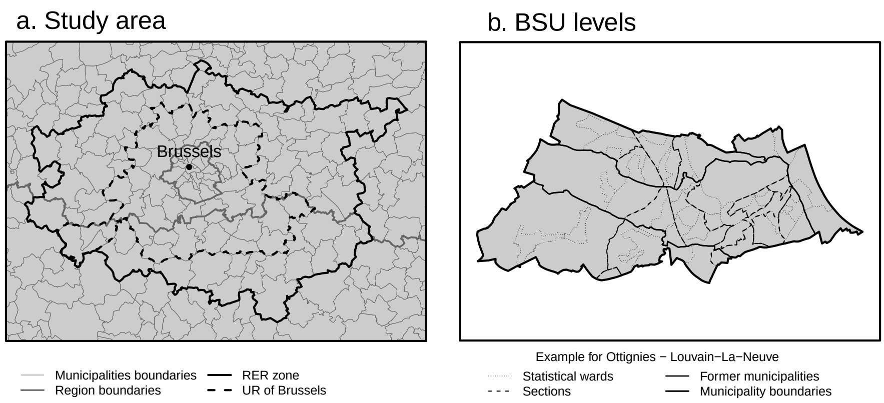

Figure C.1 – Study area and BSU levels

BSU	**Variables**	(1)	(2)	**Specifications** **(3)**	(4)	(5)
Statistical wards	POP_DENS		-0.48*** (0.09)		-0.83*** (0.07)	
(n BSU = 2074)	ACC_JOBS	-0.39** (0.09)	3.33*** (0.40)			-0.39** (0.09)
(n Jobs = 341 921)	TIME_BXL			**-0.26* (0.09)**	-0.12*** (0.02)	3.01*** (0.44)
	DIST_TRAIN	2.59*** (0.36)	-0.17*** (0.02)	**-0.69*** (0.07)**	-0.33** (0.09)	0.25 (0.15)
	DIST_HGW	-0.49*** (0.09)	-0.34* (0.09)	**-0.78*** (0.07)**	-0.70*** (0.07)	-0.49*** (0.09)
	Rho2	0,48	0,50	**0,46**	0,47	0,48
Section	POP_DENS		-0.58*** (0.10)		-0.89*** (0.09)	
(n BSU = 550)	ACC_JOBS	-0.44** (0.11)	3.06*** (0.43)			-0.44** (0.11)
(n Jobs = 341 921)	TIME_BXL			**-0.55*** (0.12)**	-0.05 (0.03)	2.84*** (0.54)
	DIST_TRAIN	2.57*** (0.39)	-0.11* (0.03)	**-0.72*** (0.09)**	-0.59*** (0.12)	0.15 (0.21)
	DIST_HGW	-0.61*** (0.10)	-0.42** (0.11)	**-0.87*** (0.09)**	-0.73*** (0.09)	-0.60*** (0.10)
	Rho2	0,52	0,53	**0,51**	0,51	0,52
Former muni.	POP_DENS		-0.64** (0.12)		-0.81*** (0.11)	
(n BSU = 173)	ACC_JOBS	-0.49* (0.14)	2.39*** (0.52)			-0.48* (0.14)
(n Jobs = 341 921)	TIME_BXL			**-1.04*** (0.15)**	0.30*** (0.06)	2.97** (0.67)
	DIST_TRAIN	3.26*** (0.46)	0.21* (0.06)	**-0.76*** (0.12)**	-0.74** (0.18)	-0.16 (0.28)
	DIST_HGW	-0.65** (0.13)	-0.55** (0.15)	**-0.97*** (0.11)**	-0.73*** (0.13)	-0.66** (0.13)
	Rho2	0,60	0,60	**0,59**	0,60	0,59
Municipalities	POP_DENS		-0.32 (0.14)		-0.59*** (0.12)	
(n BSU = 62)	ACC_JOBS	-0.41 (0.16)	5.17*** (0.76)			-0.39 (0.16)
(n Jobs = 341 921)	TIME_BXL			**-1.09*** (0.20)**	-0.09 (0.07)	1.86* (0.69)
	DIST_TRAIN	2.73*** (0.50)	-0.45*** (0.09)	**-0.59** (0.14)**	-1.20*** (0.22)	-0.53 (0.29)
	DIST_HGW	-0.39* (0.13)	-0.22 (0.16)	**-0.55*** (0.12)**	-0.62** (0.14)	-0.41* (0.13)
	Rho2	0,49	0,46	**0,45**	0,45	0,45

Table C.1 – Parameter estimates of the *Monocentric* case study (benchmark model in bold font; ns = not significant at 0,05%; between brackets = standard deviation)

BSU	Variables	Specifications (1)	(2)	**(3)**	(4)	(5)
Statistical wards	POP_DENS		-0.59*** (0.08)		-0.64*** (0.08)	
(n BSU = 4223)	ACC_JOBS	-0.33** (0.07)	0.87* (0.26)			-0.32** (0.08)
(n Jobs = 65 810)	TIME_BXL			**0.08 (0.08)**	-0.22*** (0.02)	0.96* (0.37)
	DIST_TRAIN	0.31 (0.24)	-0.24*** (0.02)	**-0.41*** (0.07)**	-0.01 (0.08)	0.34 (0.14)
	DIST_HGW	-0.46*** (0.07)	-0.32** (0.08)	**-0.49*** (0.07)**	-0.46*** (0.07)	-0.46*** (0.07)
	Rho2	0,36	0,40	**0,36**	0,39	0,36
Section	POP_DENS		-0.86*** (0.06)		-0.87*** (0.06)	
(n BSU = 1217)	ACC_JOBS	-0.35*** (0.06)	0.64* (0.19)			-0.34*** (0.06)
(n Jobs = 65 810)	TIME_BXL			**-0.009 (0.07)**	-0.11** (0.02)	0.83* (0.29)
	DIST_TRAIN	0.36 (0.18)	-0.12*** (0.02)	**-0.42*** (0.06)**	-0.09 (0.07)	0.25 (0.12)
	DIST_HGW	-0.80*** (0.06)	-0.35*** (0.06)	**-0.82*** (0.06)**	-0.44*** (0.05)	-0.80*** (0.06)
	Rho2	0,41	0,42	**0,41**	0,41	0,41
Former muni.	POP_DENS		-0.77*** (0.07)		-0.78*** (0.07)	
(n BSU = 473)	ACC_JOBS	-0.86*** (0.08)	-0.47 (0.24)			-0.86*** (0.08)
(n Jobs = 65 810)	TIME_BXL			**-0.17 (0.09)**	0.34*** (0.05)	0.40 (0.39)
	DIST_TRAIN	0.43 (0.20)	0.36*** (0.05)	**-0.89*** (0.08)**	0.16 (0.11)	-0.01 (0.18)
	DIST_HGW	-0.95*** (0.07)	-0.94*** (0.09)	**-0.95*** (0.07)**	-0.90*** (0.08)	-0.95*** (0.07)
	Rho2	0,51	0,52	**0,51**	0,52	0,51
Municipalities	POP_DENS		-0.75*** (0.09)		-0.80*** (0.09)	
(n BSU = 126)	ACC_JOBS	-0.94*** (0.10)	-2.22*** (0.33)			-0.94*** (0.10)
(n Jobs = 65 810)	TIME_BXL			**0.75*** (0.11)**	0.12 (0.06)	-0.18 (0.47)
	DIST_TRAIN	-1.36*** (0.23)	0.27** (0.07)	**-0.93*** (0.09)**	0.89*** (0.13)	0.67 (0.23)
	DIST_HGW	-0.83*** (0.08)	-0.99*** (0.10)	**-0.85*** (0.08)**	-0.91*** (0.09)	-0.85*** (0.08)
	Rho2	0,39	0,39	**0,39**	0,39	0,39

Table C.2 – Parameter estimates of the *Polycentric* case study (benchmark model in bold font; ns = not significant at 0,05%; between brackets = standard deviation)

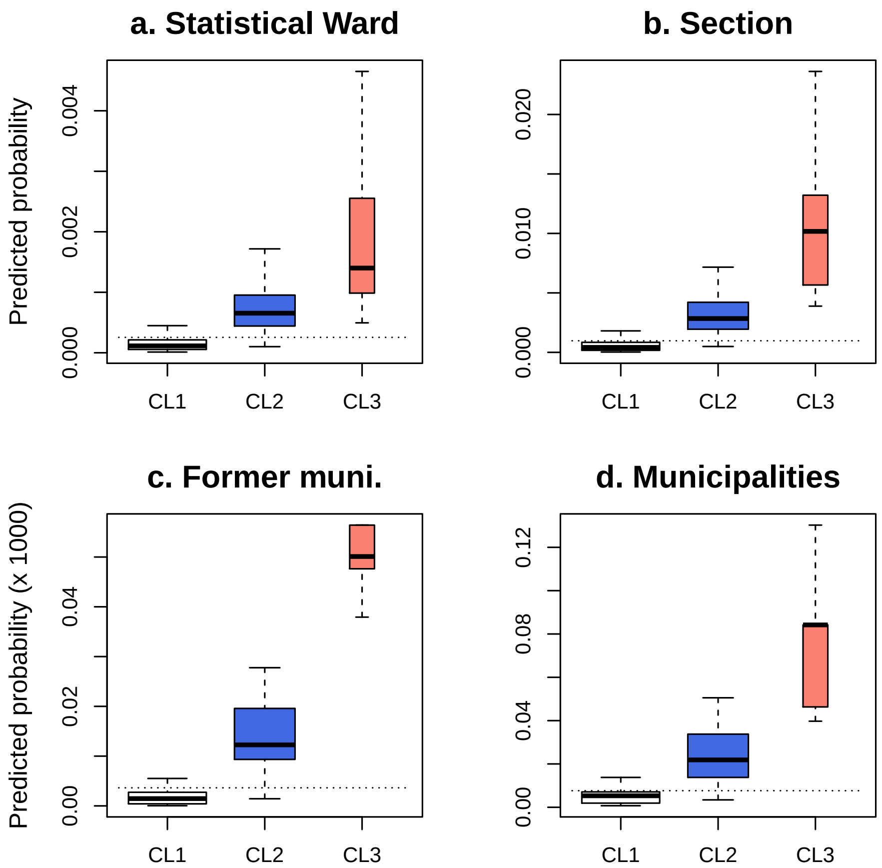

Figure C.2 – Probability of location by cluster for the *Monocentric* case study (box-plots width is function of the number of observations)

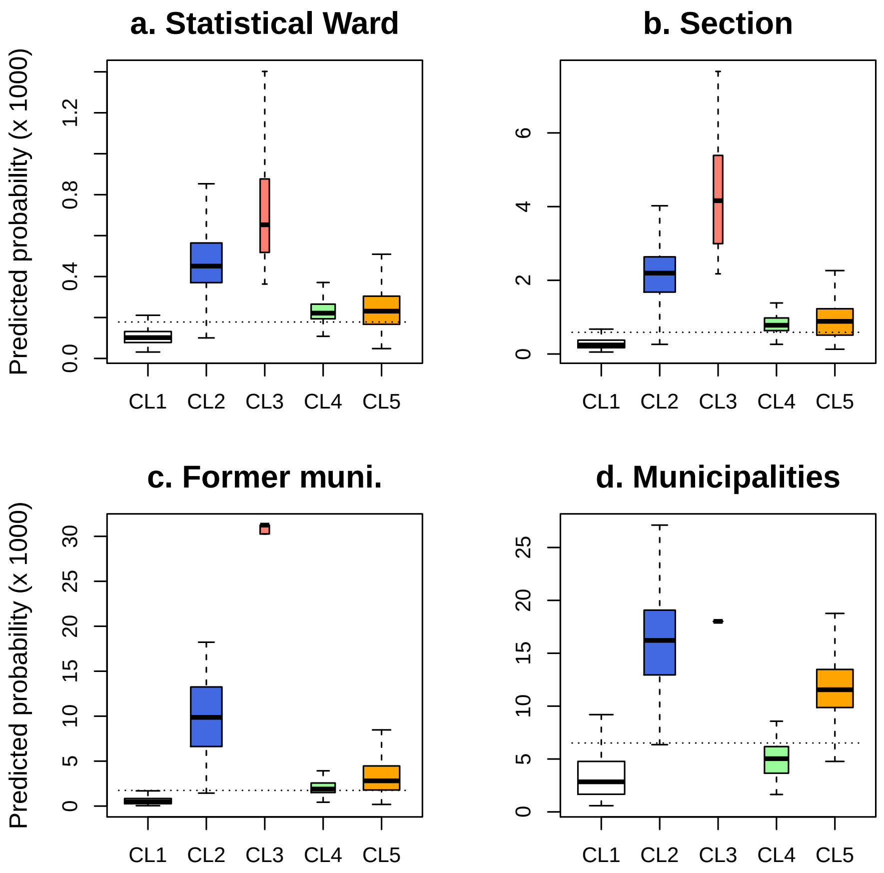

Figure C.3 – Probability of location by cluster for the *Monocentric* case study (box-plots width is function of the number of observations)

Variable	**BSU (1)**	**BSU (2)**	**p-value** *Monocentric*	**p-value** *Polycentric*
TIME_BXL	Municipalities	Former municipalities	0.05	***
		Sections	***	***
		Stat. ward	***	***
	Former municipalities	Sections	***	***
		Stat. ward	***	***
	Sections	Stat. ward	***	***
DIST_TRAIN	Municipalities	Former municipalities	***	***
		Sections	***	0.01
		Stat. ward	***	***
	Former municipalities	Sections	***	***
		Stat. ward	***	***
	Sections	Stat. ward	***	***
DIST_HGW	Municipalities	Former municipalities	***	0.001
		Sections	***	***
		Stat. ward	***	***
	Former municipalities	Sections	0.02	***
		Stat. ward	***	***
	Sections	Stat. ward	0.2	1

Table C.3 – Variations of parameter estimates through BSU levels (pairwise t-test; significance level: *** = $\alpha \leq 0.001$; Bonferroni adjustment of p-value)

BSU	**Variable**	**Spec.**	**p-value** *Monocentric*	*Polycentric*
Municipalities	TIME_BXL	(4)	0.002	***
		(5)	***	0.004
	DIST_TRAIN	(1)	***	1
		(2)	***	***
		(4)	0.9	***
		(5)	***	1
	DIST_HGW	(1)	***	1
		(2)	***	***
		(4)	1	1
		(5)	***	1
Former muni.	TIME_BXL	(4)	***	***
		(5)	***	***
	DIST_TRAIN	(1)	***	1
		(2)	***	***
		(4)	***	***
		(5)	***	1
	DIST_HGW	(1)	***	0.01
		(2)	***	***
		(4)	***	1
		(5)	***	0.02
Section	TIME_BXL	(4)	0.1	***
		(5)	***	***
	DIST_TRAIN	(1)	***	0.4
		(2)	***	***
		(4)	1	***
		(5)	***	1
	DIST_HGW	(1)	***	***
		(2)	***	***
		(4)	1	0.08
		(5)	***	***
Statistical wards	TIME_BXL	(4)	***	***
		(5)	***	***
	DIST_TRAIN	(1)	***	0.02
		(2)	***	***
		(4)	***	***
		(5)	***	***
	DIST_HGW	(1)	***	***
		(2)	***	***
		(4)	1	***
		(5)	***	***

Table C.4 – Variations of parameter estimates through specifications (pair-wise t-test; significance level: *** = $\alpha \leq 0.001$; Bonferroni adjustment of p-value)

Appendix D

Appendices of Chapter 5

D.1 Database of a zone-version of *UrbanSim*

Table	Field	Definition	Type	Value
annual employment control totals	home based status	Home-located job (yes/no)	[0 : 1]	Fixed
	number of jobs	Number of jobs	Numeric	Fixed
	sector id	Id of the activity sector	Integer	Fixed
	year	Year	Integer	Fixed
annual household	total number of households	Number of households	Numeric	Fixed
control totals	year	Year	Integer	Fixed
annual household relocation rates	age max	Upper age limit	Integer	Fixed
	age min	Lower age limit	Integer	Fixed
	income max	Upper income limit	Integer	Fixed
	income min	Lower income limit	Integer	Fixed
	probability of relocating	Annual probability of relocation	0 to 1	Fixed
annual job relocation rates	job relocation probability	Annual probability of relocation	0 to 1	Fixed
	sector id	Id of the activity sector	Integer	Fixed
building sqft per job	building sqft per job	Floor space for one job	Numeric	Fixed
	zone id	Id of the zone	Integer	Fixed
	building type id	Id of the building type	Integer	Fixed
building types	building type id	Id of the building type	Integer	Fixed

Continued on next page ...

... continuated from previous page

Table	Field	Definition	Type	Value
	building type name	Name (optional)	Character	Fixed
	is residential	Residential building (yes/no)	[0 : 1]	Fixed
buildings	building id	Id of the building	Integer	Fixed
	building type id	Id of the building type	Integer	Fixed
	zone id	Id of the zone	Integer	Fixed
	land area	Footprint of the building (optional)	Numeric	Fixed
	non residential sqft	Existing non-residential floor space	Numeric	Dynamic
	non residential sqft capacity	Maximal non-residential floor space	Numeric	Dynamic
	residential units	Existing number of residential units	Numeric	Dynamic
	residential units capacity	Maximal number of residential units	Numeric	Dynamic
	average value per unit	Monetary value of one non-residential square feet or one residential unit	Numeric	Dynamic
development event history	zone id	Id of the zone	Integer	Fixed
	building id	Id of the building	Integer	Fixed
	building type id	Id of the building type	Integer	Fixed
	change type	Addition (A); destruction (D); or renovation (R)	Character	Fixed
	land area	Affected foot print	Numeric	Fixed

Continued on next page ...

... continuated from previous page

Table	Field	Definition	Type	Value
	non residential sqft	Number of non-residential square feet affected	Numeric	Fixed
	residential units	Number of residential units affected	Integer	Fixed
	scheduled year	Year of change	Integer	Fixed
employment sectors	name	Name (optional)	Character	Fixed
	sector id	Id of the activity sector	Integer	Fixed
fazes	faz id	Id of the aggregated zone	Integer	Fixed
	faz area	Total surface	Numeric	Fixed
	xcoord	X coordinate of centroid	Numeric	Fixed
	ycoord	Y coordinate of centroid	Numeric	Fixed
home based status	home based status	Home-based job (yes/no)	[0 : 1]	Fixed
	name	Name (optional)	Character	Fixed
households	household id	Id of the household	Integer	Fixed
	building id	Id of the building	Integer	Dynamic
	age of head	Age (year) of the household's head	Numeric	Fixed
	cars	Number of cars	Integer	Fixed
	children	Presence of children	[0 : 1]	Fixed
	education high	High education level (yes/no)	[0 : 1]	Fixed
	income	Total annual income	Numeric	Fixed
	persons	Number of persons	Integer	Fixed
	workers	Number of workers	Integer	Fixed

Continued on next page ...

... continuated from previous page

Table	Field	Definition	Type	Value
jobs	job id	Id of the job	Integer	Fixed
	building id	Id of the building	Integer	Dynamic
	building type	Id of the building type	Integer	Fixed
	home based status	Home-based job (yes/no)	[0 : 1]	Fixed
	sector id	Id of the activity sector	Integer	Fixed
persons	person id	Id of the person	Integer	Fixed
	household id	Id of the household	Integer	Fixed
	job id	Id of the job	Integer	Dynamic
	origin	Id of the origin zone	Integer	Dynamic
	destination	Id of the destination zone	Integer	Dynamic
	home2work travel time min	Home to work travel time (minutes)	Numeric	Dynamic
	work2home travel time min	Work to home travel time (minutes)	Numeric	Dynamic
target vacancies	building type id	Id of the building type	Integer	Fixed
	is residential	Residential building (yes/no)	[0 : 1]	Fixed
	target vacancy	Average long-term vacancy rate	0 to 1	Fixed
	year	Year	Numeric	Fixed
travel data	from zone id	Id of the origin zone	Integer	Fixed
	to zone id	Id of the destination zone	Integer	Fixed
	am pk period drive alone vehicle trips	Number of trips (morning peak hour)	Numeric	Dynamic

Continued on next page ...

... continuated from previous page

Table	Field	Definition	Type	Value
	am single vehicle to work travel time	Travel time by car (morning peak hour)	Numeric	Dynamic
	bike time in minutes	Travel time by bike	Numeric	Fixed
	single vehicle to work travel cost	Travel cost by car (morning peak hour)	Numeric	Dynamic
	vehicle free speed travel time	Travel time by car (without congestion)	Numeric	Fixed
	walk time in minutes	Travel time by foot walk	Numeric	Fixed
zones	zone id	Id of the zone	Integer	Fixed
	faz id	Id of the aggregated zone	Integer	Fixed
	d cbd	Euclidean distance to the CBD	Numeric	Fixed
	car accessibility	Car accessibility to jobs (congestion included)	Numeric	Dynamic
	freespeed accessibility	Car accessibility to jobs (without congestion)	Numeric	Dynamic
	pt accessibility	Public transport accessibility to jobs	Numeric	Dynamic
	walk accessibility	Foot walk accessibility to jobs	Numeric	Dynamic
	surface	Total surface	Numeric	Fixed
	xcoord	X coordinate of centroid	Numeric	Fixed
	ycoord	Y coordinate of centroid	Numeric	Fixed

Table D.1 – Database of *UrbanSim* (OPUS v4.3, zone-version)

D.2 Sub models of a zone-version of *UrbanSim*

Submodel	Type	Action	Depends on	Dependencies
development project transition model (DPTM)	Determinist	Computes the need for development projects of new non-residential floor space or residential units, as (existing - occupied) versus target vacancy rate	target vacancies, buildings, households, jobs, zones	
residential development project location choice model (RDPLCM)	DCM	Determines the location of the residential development project. One specification per residential building type.	development event history, buildings, households, jobs, zones, outputs from DPTM	buildings
non residential development project location choice model (NRD-PLCM)	DCM	Determines the location of the non-residential development project. One specification per non-residential building type.	development event history, buildings, households, jobs, zones, outputs from DPTM	buildings
add projects to buildings (APB)	Determinist	Update the required fields in the buildings table according to the development projects	outputs from RD-PLCM and NRD-PLCM	buildings
household transition model (HTM)	Sampling	Draws a pool of new households by random sampling of existing households	annual household control totals, households	

Continued on next page ...

... continuated from previous page

Submodel	Type	Action	Depends on	Dependencies
household relocation model (HRM)	Sampling	Draws a pool of relocating households by weighted sampling of existing households	annual household relocation rates, households	
household location choice model (HLCM)	DCM	Determines the future residential location of the new + relocating households	outputs from HTM and HRM, buildings, households, jobs, zones	households
employment transition model (ETM)	Sampling	Draws a pool of new jobs by random sampling of existing job, by activity sector	annual employment control totals, jobs	
employment relocation model (ERM)	Sampling	Draws a pool of relocating jobs by weighted sampling of existing jobs, by activity sector	annual job relocation rates, jobs	
employment location choice model (ELCM)	DCM	Determines the future location of the new + relocating jobs. One specification for home-based jobs and one per activity sector for non-home-based jobs	outputs from ETM and ERM, buildings, households, jobs, zones	jobs

Continued on next page ...

... continuated from previous page

Submodel	**Type**	**Action**	**Depends on**	**Dependencies**
real estate price model (REPM)	Regression	Update the average value per unit field from the building table. One specification per building type	buildings, households, jobs, zones	buildings
travel model (TM)	Determinist	Flag for calling the external travel model (here *MATsim*)	zones, households, persons, jobs, travel data	zones, persons, travel data

Table D.2 – Sub models of *UrbanSim* (OPUS v4.3, zone-version).

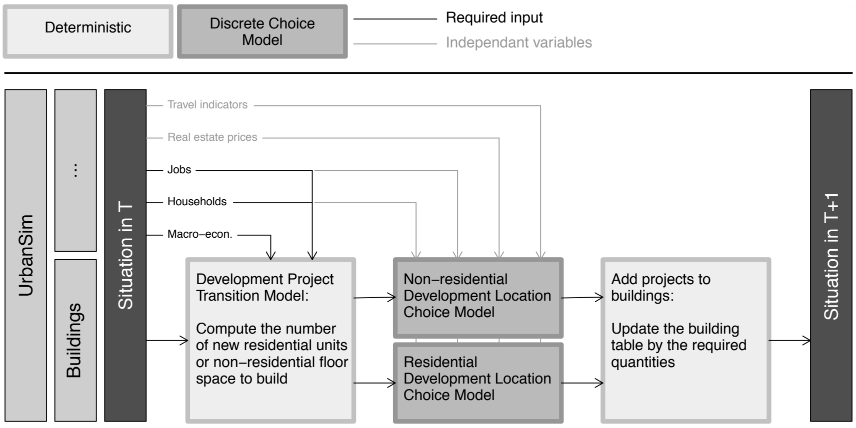

Figure D.1 – The buildings sequence

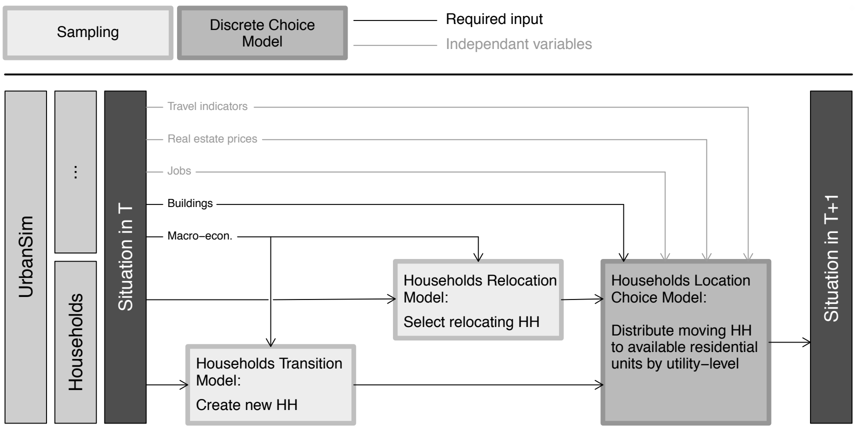

Figure D.2 – The households sequence

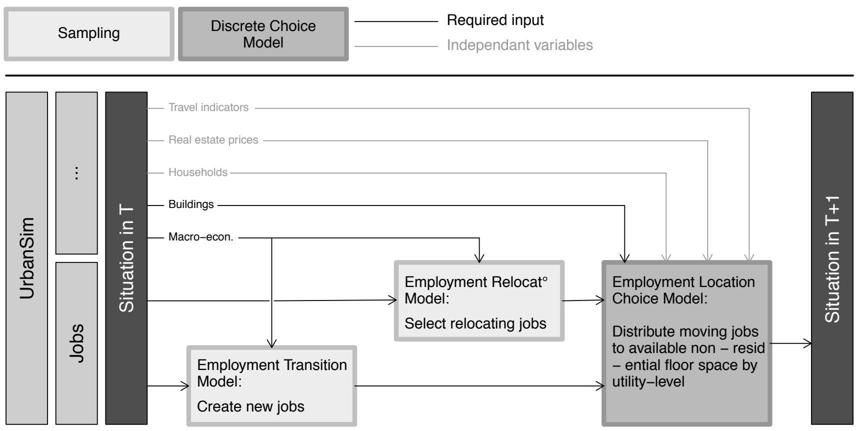

Figure D.3 – The jobs sequence

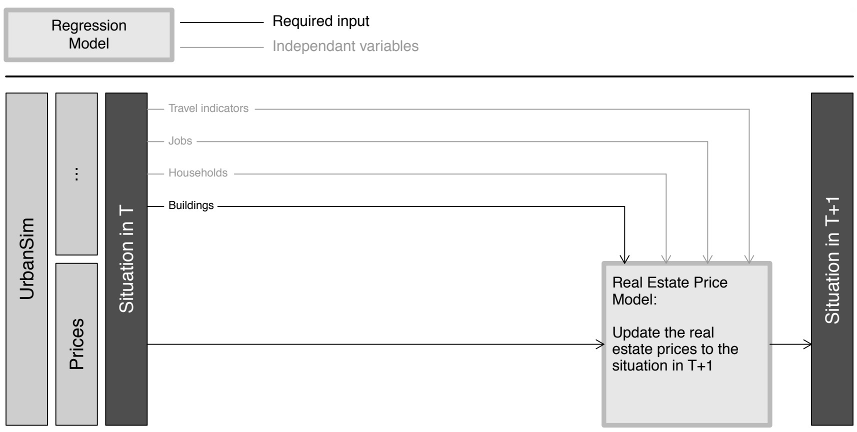

Figure D.4 – The real estate prices sequence

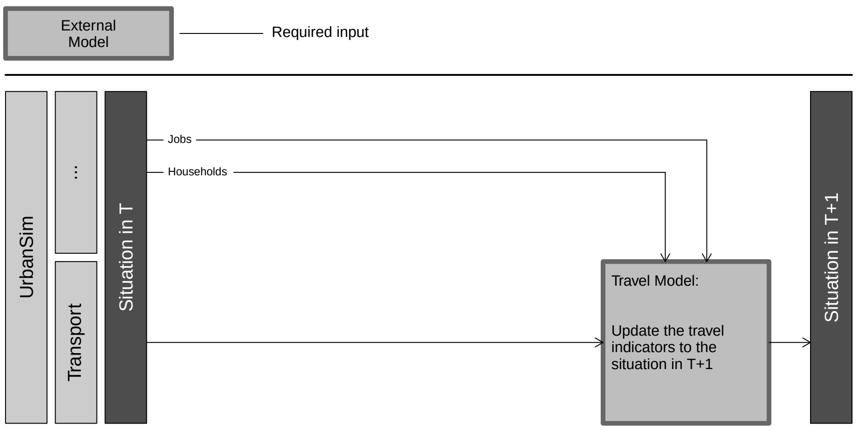

Figure D.5 – The external travel model sequence

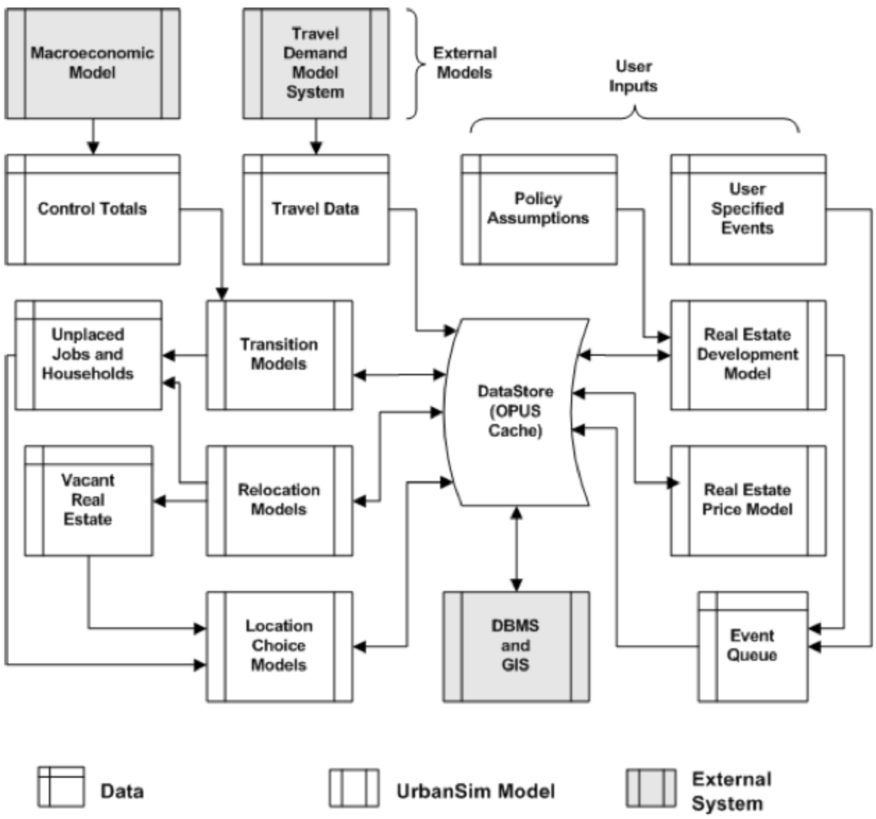

Figure D.6 – Workflow of *UrbanSim* (figure from Waddell, 2000)

D.3 Additional tables and figures

Step	Description	Inputs	Outputs
Import	Read user-defined parameters from the indicated csv file	Path to the csv file	
General	Produce "general" tables		BUILDING TYPES; HOME BASED STATUS; EMPLOYMENT SECTOR; BUILDING SQFT PER JOB
Macro	Produce tables summarising macro economic assumptions	Number of households in t_0; Population growth and relocation rate, target vacancy rate	ANNUAL HOUSEHOLD CONTROL TOTALS; ANNUAL HOUSEHOLD RELOCATION RATES; ANNUAL EMPLOYMENT CONTROL TOTALS; ANNUAL JOB RELOCATION RATES; TARGET VACANCIES
Grid	Create the grid of the synthetic city	Number of rows and columns; surface of the zones; location and relative size of the CBDs	Euclidean distance to the CBDs; ZONES; TRAVEL DATA
Potentials	Computation of the potentials of each zone, for households and non-home-based jobs	Distance to the CBD(s); distance-decay parameters	$P_i(h)$; $P_i(j)$

Continued on next page ...

... continuated from previous page

Step	Description	Inputs	Outputs
Households	Generation of households	Number of households in t_0; $P_i(h)$	HOUSEHOLDS
Jobs	Generation of jobs	Number of households in t_0; $P_i(j)$	JOBS
Persons	Creation of persons	HOUSEHOLDS; JOBS	PERSONS
Buildings	Generation of buildings	ZONES; HOUSEHOLDS; JOBS BUILDING SQFT PER JOB; TARGET VACANCIES	BUILDINGS;
Development events	Creation of (past) development events	BUILDINGS	DEVELOPMENT EVENT HISTORY
Network	Creation of the *MATsim* network	ZONES	xml network file
Export	Writes the tables as csv files	All tables	Path to the export directory

Table D.3 – Generation of the synthetic city (outputs in SMALL CAPITALS indicate the *UrbanSim* tables)

Extent	**Variable**		**Case study**	
		Equal-size CBDs	Small W. CBD	Large W. CBD
Complete	P_h (10^{-6})	-4.26* (2.07)	-0.39 (0.37)	-1.63*** (0.37)
	A_i	0.053 (0.03)	0.07*** (0.02)	0.008 (0.021)
	D_{cbd}	-0.005 (0.01)	0.02** (0.006)	-0.007 (0.006)
	AIC	106 577	106 473	1064 71
	LL_{ratio}	0.0002	0.0001	0.0002
E50	P_h (10^{-6})	-2.28*** (0.54)	-0.71 (0.58)	-0.61 (0.36)
	A_i	0.022* (0.011)	0.045*** (0.013)	0.006 (0.005)
	D_{cbd}	-3.21e-06 (0.0007)	-0.001 (0.001)	-0.002 (0.001)
	AIC	88 683	82 967	94 288
	LL_{ratio}	0.0002	0.0002	0.0002
E45	P_h (10^{-6})	-1.91** (0.59)	-1.57 (1.30)	
	A_i	0.006 (0.012)	0.01 (0.03)	0.003 (0.005)
	D_{cbd}	0.002 (0.001)	0.004 (0.002)	0.0007 (0.001)
	AIC	75 397	65 394	85 413
	LL ratio	0.0002	0.0002	0.0002
E40	P_h (10^{-6})	-2.67*** (0.66)	-3.06** (1.08)	-1.14 (1.10)
	A_i	0.012 (0.014)	0.008 (0.03)	0.015 (0.024)
	D_{cbd}	-0.0007 (0.003)	0.0006 (0.005)	-0.002 (0.003)
	AIC	64 195	50 620	77 736
	LL_{ratio}	0.0002	0.0004	0.0003
E35	P_h (10^{-6})	-1.0 (1.4)	-10.1*** (1.6)	-3.35** (1.05)
	A_i	0.003 (0.045)	0.13*** (0.03)	0.008 (0.006)
	D_{cbd}	-0.006 (0.01)	0.0009 (0.007)	-0.023** (0.008)
	AIC	57 231	41 679	72 730
	LL_{ratio}	0.0002	0.0004	0.0003
E30	P_h (10^{-6})	-2.89 (2.83)	-0.71 (3.86)	-0.65 (2.12)
	A_i	0.05 (0.04)	0.013 (0.034)	0.001 (0.006)
	D_{cbd}	0.001 (0.02)	0.014 (0.017)	-0.015 (0.019)
	AIC	53 287	37 136	69 320
	LL_{ratio}	0.0002	0.0006	0.0002
E25	P_h (10^{-6})	-1.20 (4.63)	-7.81 (7.89)	-6.3* (3.1)
	A_i	0.11* (0.044)	0.006 (0.03)	0.0004 (0.007)
	D_{cbd}	0.034 (0.035)	-0.026 (0.038)	-0.06* (0.03)
	AIC	49 338	33 725	64 780
	LL_{ratio}	0.0003	0.0006	0.0002

Table D.4 – Households location choices sub model (with P_h the price of houses; A_I the car accessibility to jobs in i; D_{cbd} the Euclidean distance to the CBD; AIC the Akaike Information Critetion; and LL_{ratio} the Log-Likelihood ratio)

Extent	Variable	Case study		
		Equal-size CBDs	Small W. CBD	Large W. CBD
Complete	P_o (10^{-7})	-11.2*** (0.07)	-0.003 (0.05)	27.8*** (0.03)
	A_i	1.82*** (0.01)	1.38*** (0.009)	0.96*** (0.007)
	AIC	110 236	113 000	117 307
	LL_{ratio}	0.46	0.45	0.43
E50	P_o (10^{-7})	50.6*** (0.03)	30.7*** (0.03)	69.4*** (0.02)
	A_i	0.58*** (0.004)	0.86*** (0.005)	0.08*** (0.002)
	AIC	125 750	120 787	132 646
	LL_{ratio}	0.37	0.38	0.34
E45	P_o (10^{-7})	80.5*** (0.05)	30.3*** (1.33)	77.4*** (0.03)
	A_i	0.22*** (0.005)	1.26*** (0.01)	-0.04*** (0.002)
	AIC	69 470	47 711	86 513
	LL_{ratio}	0.37	0.39	0.39
	Obs.	16 287	11 589	20 922
E40	P_o (10^{-7})	65.5*** (0.05)	2.81*** (1.36)	5.91*** (0.05)
	A_i	0.38*** (0.006)	1.14*** (0.01)	1.42*** (0.01)
	AIC	63 575	40 042	74 505
	LL_{ratio}	0.38	0.42	0.45
E35	P_o (10^{-7})	-10.1*** (1.0)	24.6*** (1.36)	70.1*** (0.03)
	A_i	1.80*** (0.02)	1.21*** (0.01)	-0.029*** (0.002)
	AIC	54 789	40 449	86 061
	LL_{ratio}	0.47	0.41	0.37
E30	P_o (10^{-7})	-14.4*** (1.03)	29.3*** (1.29)	69.2*** (0.03)
	A_i	1.88*** (0.02)	1.16*** (0.01)	-0.008** (0.003)
	AIC	54 389	40 104	86 060
	LL_{ratio}	0.47	0.42	0.37
E25	P_o (10^{-7})	-1.92* (0.88)	30.0*** (1.3)	69.2*** (0.03)
	A_i	1.59*** (0.01)	1.14*** (0.01)	-0.012*** (0.003)
	AIC	56 582	40 141	86 126
	LL_{ratio}	0.45	0.42	0.37

Table D.5 – Employment location choices sub model (with P_o the price of offices; A_I the car accessibility to jobs in i; AIC the Akaike Information Critetion; and LL_{ratio} the Log-Likelihood ratio)

Extent	Variable	Real estate prices		
		Equal-size CBDs	Small W. CBD	Large W. CBD
Complete	Constant	6.21 (0.08)	5.61 (0.06)	5.60 (0.06)
	Job density	0.35 (0.004)	0.32 (0.004)	0.32 (0.004)
	Pop. Density	0.39 (0.01)	0.51 (0.01)	0.51 (0.01)
	Adj. R^2	0.95	0.95	0.95
	Obs.	1 500	1 500	1 500
E50	Constant	6.27 (0.08)	5.71 (0.07)	5.72 (0.07)
	Job density	0.36 (0.005)	0.34 (0.005)	0.33 (0.004)
	Pop. Density	0.38 (0.01)	0.48 (0.01)	0.49 (0.01)
	Adj. R^2	0.95	0.95	0.95
	Obs.	1 250	1 250	1 250
E45	Constant	5.87 (0.08)	5.52 (0.07)	5.41 (0.07)
	Job density	0.32 (0.005)	0.30 (0.004)	0.31 (0.005)
	Pop. Density	0.46 (0.01)	0.53 (0.01)	0.55 (0.01)
	Adj. R^2	0.95	0.95	0.96
	Obs.	1 125	1 125	1 125
E40	Constant	5.86 (0.08)	5.54 (0.08)	5.62 (0.07)
	Job density	0.33 (0.005)	0.31 (0.005)	0.33 (0.005)
	Pop. Density	0.46 (0.01)	0.52 (0.01)	0.51 (0.01)
	Adj. R^2	0.96	0.94	0.96
	Obs.	1 000	1 000	1 000
E35	Constant	5.98 (0.09)	6.14 (0.1)	5.68 (0.07)
	Job density	0.34 (0.006)	0.34 (0.006)	0.33 (0.005)
	Pop. Density	0.44 (0.01)	0.39 (0.02)	0.50 (0.01)
	Adj. R^2	0.96	0.95	0.96
	Obs.	875	875	875
E30	Constant	6.21 (0.11)	6.38 (0.12)	6.10 (0.11)
	Job density	0.35 (0.006)	0.35 (0.006)	0.34 (0.006)
	Pop. Density	0.39 (0.02)	0.34 (0.02)	0.42 (0.02)
	Adj. R^2	0.95	0.95	0.96
	Obs.	750	750	750
E25	Constant	6.94 (0.16)	6.93 (0.16)	7.13 (0.16)
	Job density	0.37 (0.007)	0.37 (0.007)	0.38 (0.007)
	Pop. Density	0.26 (0.03)	0.23 (0.03)	0.24 (0.03)
	Adj. R^2	0.94	0.94	0.94
	Obs.	625	625	625

Table D.6 – Non-residential buildings real estate price sub model (population and job density expressed in log; all parameters significant for $\alpha \leq 0.001$)

Indicator	**Extent**	**Equal size CBD**			**Small west CBD**			**Large west CBD**		
		West CBD	Suburban	East CBD	West CBD	Suburban	East CBD	West CBD	Suburban	East CBD
Households	*Complete*	0.46	0.07	0.46	0.31	0.07	0.61	0.61	0.07	0.31
	E50	0.002	-0.03	0.002	-0.009	-0.04	0.01	0.001	0.03	-0.02
	E45	-0.001	-0.007	0.007	-0.005	-0.03	0.01	0.002	-0.0001	-0.01
	E40	0.002	0.0006	-0.02	0.005	-0.03	0.003	0.001	-0.005	-0.01
	E35	0.006	-0.04	/	0.01	-0.06	/	0.008	-0.07	/
	E30	0.003	-0.04	/	-0.005	0.05	/	0.007	-0.11	/
	E25	0	/	/	0	/	/	0	/	/
Jobs	*Complete*	0.50	0	0.50	0.35	0	0.65	0.65	0	0.35
	E50	-2.11	0	2.26	-12.71	0	7.33	2.96	0	-5.94
	E45	0.77	0	-11.14	-0.02	0	0.16	0.23	0	-6.53
	E40	0.04	0	-59.63	-0.001	0	0.77	-0.003	0	10.69
	E35	0	0	/	0	0	/	0	0	/
	E30	0	0	/	0	0	/	0	0	/
	E25	0	/	/	0	/	/	0	/	/
Price	*Complete*	24 757	4 499	24 760	16 698	4 541	33 209	33 211	4 539	16 705
	E50	-1.09	-0.01	-1.16	-5.24	-0.24	-5.68	-2.56	-0.05	-2.75
	E45	-0.45	0.38	-0.23	11.91	-0.11	8.10	3.99	0.11	2.54
	E40	-0.43	0.34	0.36	8.38	0.31	1.53	2.45	-0.06	0.17
	E35	-0.65	0.55	/	-0.70	-0.62	/	-1.73	-0.14	/
	E30	-1.09	0.13	/	-1.09	-1.13	/	-2.39	-0.52	/
	E25	-2.09	/	/	-1.95	/	/	-3.87	/	/
Commuting time	*Complete*	46.88	64.15	45.43	35.49	79.96	77.75	62.51	68.19	34.65
	E50	-6.75	-16.12	-30.91	-7.02	-35.65	-49.00	-10.10	-8.89	-15.75
	E45	-7.36	-7.37	7.57	-17.38	-47.29	-54.01	-8.76	12.84	53.28
	E40	-9.62	-3.28	18.98	-18.52	-40.05	-34.08	-9.01	15.05	75.07
	E35	-12.84	-4.67	/	-21.14	-41.94	/	-9.11	14.51	/
	E30	-16.29	-11.97	/	-23.00	-39.51	/	-13.51	-3.37	/
	E25	-20.67	/	/	-26.56	/	/	-16.56	/	/

Table D.7 – Variations per macro zones (values = t_{20} for the *Complete* extent; relative differences as $\frac{Complete-E_x}{Complete}$, in %, for *Small extents*)

APPENDIX E

Appendices of Chapter 6

E.1 Additional tables and figures

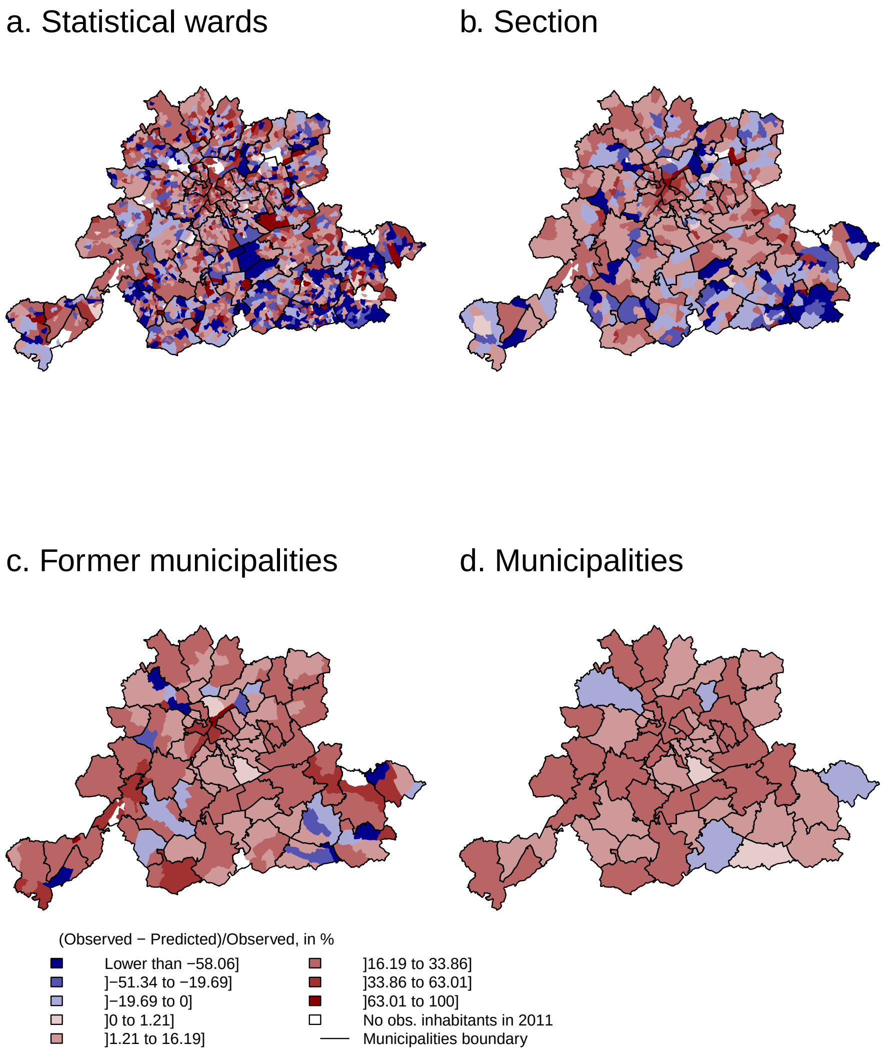

Figure E.1 – **Calibration** (relative variations of the inhab/km^2; observed data from the 2011 population census; predicted data from the *Baseline* scenario)

Category	Variable	Type	Level	REPM	ELCM	HLCM
Agglomeration economies	(Log of) Pop. Density	Endogenous	Zones	Yes	Yes	Yes
	(Log of) Job Density (total)	Endogenous	Zones	Yes	Yes	
	- *in Industrial activities*				Yes	
	- *in Services*				Yes	
	- *in Retail*				Yes	
	- *in Hotel/Restaurant/Bar*				Yes	
	- *in Government/Public sector*				Yes	
	- *in Education*				Yes	
	- *in Health*				Yes	
	- *in Leisure activities*				Yes	
Accessibility amenities	Car Accessibility to Jobs	MATsim	Zones	Yes	Yes	Yes
	(Log of) Dist. to CBD	Fixed	Zones	Yes	Yes	Yes
Socio-economic amenities	High income Households	Endogenous	Zones	Yes	Yes	
	Low income Households	Endogenous	Zones		Yes	
	HH with univ. degree holder	Endogenous	Zones			Yes
Local amenities	(Log of) Houses prices	Endogenous	Municipalities	Dependant	Yes	Yes
	Local taxes	Fixed	Municipalities	Yes		
	Green Amenities score	Fixed	Zones	Yes	Yes	

Table E.1 – Specifications of the econometric sub models (REPM = Real estate price sub model; ELCM = Employment location choice sub model; HLCM = Households location choice sub model)

Variable	Units	Statistical wards level Min.	Mean	Max.	SD
(Log of) Pop. Density	Hab/km^2	0.00	8.16	10.75	1.52
(Log of) Job Density (total)	Job/km^2	0.00	7.75	11.26	1.97
- in Industrial activities		0.00	8.03	12.22	1.67
- in Services		0.00	8.84	13.70	1.58
- in Retail		0.00	7.28	11.84	1.34
- in Hotel/Restaurant/Bar		0.00	7.11	13.17	1.49
- in Government/Public sector		0.00	8.63	14.12	1.71
- in Education		0.00	6.99	11.25	1.30
- in Health		0.00	7.08	11.76	1.29
- in Leisure activities		0.00	6.22	11.03	1.25
Car Accessibility to Jobs	Logsum	-3.72	7.41	11.25	2.22
(Log of) Dist. to CBD	Meters	4.54	9.44	10.61	0.84
High income Households	%	0.00	11.33	100	8.53
Low income Households	%	0.00	6.64	100	5.97
HH with univ. degree holder	%	0.00	9.01	100	7.27
(Log of) Houses prices	1 000 €	11.37	11.68	11.92	0.10
Local taxes	%	4.00	6.25	8.00	0.78
Green Amenities score	0 to 1	0.00	0.60	1.00	0.18

Table E.2 – Variables of the econometric sub models (at the Statistical wards level; SD = Standard deviation)

Building type	Variable	BSU Stat. wards	Section	Former muni.	Municipalities
Houses	(Log of) Pop. density		-0.0009 (0.0007)	0.024*** (0.002)	0.051*** (0.005)
	Car accessibility to Jobs	0.009*** (0.0006)	0.013*** (0.001)		
	(Log of) Dist. to CBD	-0.024*** (0.001)			
	Local taxes	-0.041*** (0.001)	-0.04*** (0.003)	-0.059*** (0.004)	-0.037*** (0.008)
	Green Amenities score	0.056*** (0.002)	0.048*** (0.006)	0.05*** (0.01)	0.12*** (0.02)
	Adj. R^2	0.21	0.19	0.34	0.39
	Observations	6 222	1 650	519	186
Flats	Car accessibility to Jobs	0.01*** (0.002)			
	(Log of) Dist. to CBD	0.030*** (0.006)		0.02 (0.02)	
	Local taxes	-0.09*** (0.006)	-0.08*** (0.01)	-0.08*** (0.01)	-0.08* (0.03)
	Green Amenities score	0.06*** (0.01)	0.06** (0.02)	0.04 (0.04)	0.09 (0.07)
	Adj. R^2	0.1 6	0.13	0.16	0.13
	Observations	2 074	550	173	62

Table E.3 – Real estate prices sub model (between brackets: standard deviation; significance level: * = $\alpha \leq 0.05$, ** = $\alpha \leq 0.01$, *** = $\alpha \leq 0.001$)

	BSU			
Variable	Stat. wards	Section	Former muni.	Municipalities
(Log of) Pop. density	0.06*** (0.01)	0.09*** (0.01)	0.82*** (0.01)	-0.05* (0.02)
(Log of) Job density	-0.09*** (0.02)	0.002 (0.01)	0.09*** (0.01)	0.55*** (0.02)
Car accessibility to Jobs	0.064*** (0.006)	0.135*** (0.007)	0.051*** (0.009)	-0.004 (0.01)
(Log of) Dist. to CBD	-0.41*** (0.01)	-0.52*** (0.01)	0.43*** (0.02)	0.34*** (0.03)
% High inc. HH × High inc. HH	0.32*** (0.02)	0.24*** (0.02)	0.38*** (0.03)	0.42*** (0.04)
% Low inc. HH × High inc. HH	0.08*** (0.007)	0.088*** (0.008)	0.08*** (0.01)	0.06*** (0.01)
(Log of) Houses price	-1.59*** (0.04)	-1.91*** (0.05)	-1.61*** (0.05)	-1.95*** (0.05)
× High inc. HH	0.51** (0.16)	0.85*** (0.16)	0.76*** (0.18)	0.60*** (0.18)
× Low inc. HH	-0.58*** (0.16)	-0.75*** (0.17)	-0.82*** (0.20)	-0.82*** (0.18)
Green Amenities score	0.14*** (0.02)	0.09*** (0.02)	0.31*** (0.02)	0.12*** (0.02)
Adj. Likelihood index	0.07	0.11	0.22	0.10
AIC	56 900	54 170	47 490	54 930

Table E.4 – Households location choice sub model (between brackets: standard deviation; significance level: * = $\alpha \leq 0.05$, ** = $\alpha \leq 0.01$, *** = $\alpha \leq 0.001$)

Activity sector	Variable	BSU Stat. wards	Section	Former muni.	Municipalities
Industrial activities	(Log) pop. density	-0.12*** (0.007)	-0.25*** (0.02)	-0.001 (0.04)	-1.02*** (0.046)
	(Log) job density	-0.68*** (0.02)	-0.41*** (0.05)	0.64*** (0.04)	0.72*** (0.08)
	- *in Industry*	1.17*** (0.03)	1.03*** (0.04)		0.62*** (0.05)
	- *in Education*			-0.21*** (0.03)	
	Car access. to Jobs	0.17*** (0.005)	0.098*** (0.021)	0.092*** (0.026)	-0.062 (0.03)
	(Log) dist. to CBD	-0.23*** (0.01)	-0.26*** (0.042)		0.15 (0.09)
	(Log) houses prices	9.3e-05*** (1.0e-05)	1.3e-04*** (3.5e-05)	0.0003*** (4.3e-05)	8.9e-05 (5.0e-05)
	% High inc. HH	-0.02*** (0.002)	-0.042** (0.014)		-0.02 (0.04)
	% High edu. HH	-0.12*** (0.005)	-0.06 (0.03)	-0.19*** (0.03)	-0.02 (0.11)
	Adj. LL Index	0.19	0.26	0.21	0.14
	AIC	122 000	11 010	11 950	12 860
Services	(Log) pop. density	-0.14*** (0.009)	-0.21*** (0.02)	-0.04 (0.02)	-1.28*** (0.04)
	(Log) job density	-0.71*** (0.03)	-0.52*** (0.05)		0.61*** (0.12)
	- *in Services*	1.24*** (0.03)	1.03*** (0.04)	0.73*** (0.02)	1.08*** (0.11)
	Car access. to Jobs	0.35*** (0.004)	0.17*** (0.02)	0.02 (0.03)	-0.18*** (0.04)
	(Log) dist. to CBD	-0.61*** (0.007)	-0.53*** (0.03)		0.17* (0.079)
	(Log) houses prices	0.0002*** (8.8e-06)	0.0001*** (3.1e-05)	0.0002*** (2.6e-05)	0.0001** (4.4e-05)
	% High inc. HH	0.002 (0.002)	-0.04*** (0.01)		-0.17*** (0.04)
	% High edu. HH	-0.08*** (0.005)	0.042 (0.035)		0.37*** (0.11)
	Adj. LL Index	0.22	0.29	0.33	0.23
	AIC	148 800	13 410	12 680	14 540
Retail activities	(Log) pop. density	-0.21*** (0.02)	-0.28*** (0.06)		-0.97*** (0.11)
	(Log) job density	-0.43*** (0.01)	-0.247*** (0.06)	-0.32*** (0.07)	0.68*** (0.09)
	- *in Retail*	1.14*** (0.05)	1.00*** (0.07)	1.08*** (0.08)	
	- *in Health*				0.34*** (0.084)

Continued on next page ...

... continuated from previous page

Activity sector	Variable	BSU Stat. wards	Section	Former muni.	Municipalities
	Car access. to Jobs	0.29*** (0.008)	0.15*** (0.04)	0.07 (0.05)	0.08 (0.07)
	(Log) dist. to CBD		-0.16* (0.06)		
	% High inc. HH		-0.02 (0.03)	0.03 (0.05)	
	% High edu. HH	-0.09*** (0.01)	-0.07 (0.09)	-0.03 (0.14)	-0.09 (0.06)
	(Log) houses prices	-2.5e-05 (1.7e-05)	-9.5e-06 (6.5e-05)	4.2e-05 (7.3e-05)	1.1e-05 (8.5e-05)
	Adj. LL Index	0.08	0.17	0.21	0.07
	AIC	36 960	3 337	3 260	3 641
Hotels, Restaurant, and Bars	(Log) pop. density	-0.09*** (0.02)	-0.08 (0.06)	-0.04 (0.11)	-1.72*** (0.14)
	(Log) job density	-0.27** (0.04)	-0.34*** (0.09)	0.02 (0.14)	1.68*** (0.13)
	- *in HORECA*	0.98*** (0.05)	0.92*** (0.09)	0.77*** (0.11)	
	- *in Health*				0.41*** (0.11)
	Car access. to Jobs	0.19*** (0.01)	0.18*** (0.05)	0.08 (0.07)	-0.13 (0.09)
	(Log) dist. to CBD	-0.79*** (0.02)	-0.52*** (0.08)	0.17 (0.16)	
	% High inc. HH	-0.001 (0.006)	0.003 (0.036)	-0.05 (0.07)	
	% High edu. HH	0.02 (0.01)	-0.02 (0.1)	0.11 (0.20)	0.22* (0.1)
	(Log) houses prices	0.0001*** (2.7e-05)	-2.4e-05 (9.1e-05)	0.0002 (0.0001)	-8.5e-05 (0.0001)
	Adj. LL Index	0.19	0.29	0.32	0.21
	AIC	19 180	1 716	1 604	1 847
Government and Public services	(Log) pop. density	-0.03 (0.01)	-0.55*** (0.04)		-1.50*** (0.05)
	(log) job density	-0.64*** (0.03)	0.31*** (0.03)	1.04*** (0.04)	0.95*** (0.12)
	- *in Retail*		0.537*** (0.04)		
	- *in Public*	1.17*** (0.04)			0.91*** (0.06)
	- *in Leisure*			0.11*** (0.03)	
	Car access. to Jobs	0.35*** (0.005)	0.58*** (0.02)	-0.09** (0.03)	-0.09 (0.05)
	(Log) dist. to CBD	-0.84*** (0.01)		0.12 (0.07)	0.12 (0.11)

Continued on next page ...

... continuated from previous page

Activity sector	Variable	BSU Stat. wards	Section	Former muni.	Municipalities
	% High inc. HH	0.006** (0.002)			-0.26*** (0.05)
	% High edu. HH	-0.06*** (0.006)	-0.27*** (0.02)		0.71*** (0.13)
	(Log) houses prices	9.2e-05*** (1.4e-05)	5.2e-05 (4.7e-05)		5.5e-05 (6.5e-05)
	Adj. LL Index	0.26	0.29	0.44	0.39
	AIC	84 720	8 150	6 390	7 021
Education	(Log) pop. density	-0.17*** (0.03)	-0.15* (0.06)	-0.15 (0.08)	-1.23*** (0.07)
	(Log) job density	-0.61*** (0.07)	-0.01 (0.06)	-0.25** (0.08)	
	- *in Retail*		0.44*** (0.06)		
	- *in Education*	1.36*** (0.08)		1.06*** (0.07)	1.51*** (0.07)
	Car access. to Jobs	0.25*** (0.01)	0.51*** (0.04)	0.03 (0.05)	0.03 (0.07)
	(Log) dist. to CBD	-0.32*** (0.02)			
	% High inc. HH	0.003 (0.0047)			
	% High edu. HH	-0.03** (0.01)	-0.04 (0.03)	-0.01 (0.04)	
	(Log) houses prices	0.0001*** (1.7e-05)	0.0001* (6.2e-05)	0.0001* (7.6e-05)	-7.4e-05 (5.5e-05)
	Adj. Log-Likelihood Index	0.10	0.15	0.32	0.22
	AIC	35 370	3 331	2 673	3 078
Health	(Log) pop. density	-0.24*** (0.02)	-0.22*** (0.05)		-1.51*** (0.09)
	(Log) job density	-0.51*** (0.04)	0.27*** (0.04)	0.22*** (0.06)	0.53*** (0.08)
	- *in Retail*		0.20*** (0.05)		
	- *in Health*	1.29*** (0.05)			1.14*** (0.08)
	- *in Leisure*			0.56*** (0.04)	
	Car access. to Jobs	0.07*** (0.008)	0.23*** (0.03)	0.13** (0.05)	0.003 (0.059)
	(Log) dist. to CBD	-0.53*** (0.02)		0.43*** (0.09)	
	% High edu. HH	0.0007 (0.01)	-0.18*** (0.03)		-0.03 (0.05)

Continued on next page ...

... continuated from previous page

Activity sector	Variable	BSU			
		Stat. wards	Section	Former muni.	Municipalities
	% High inc. HH	-0.01*** (0.003)			
	(Log) houses prices	0.0003*** (1.4e-05)	0.0003*** (5.1e-05)		0.0001 (7.4e-05)
	Adj. LL Index	0.11	0.15	0.25	0.16
	AIC	46 370	4 468	3 824	4 494
Leisure activities	(Log) pop. density	-0.03 (0.03)			-0.91*** (0.14)
	(Log) job density	-0.31*** (0.04)	-0.43*** (0.05)		1.81*** (0.18)
	- *in Health*				0.12 (0.15)
	- *in Leisure*	1.05*** (0.05)	1.01*** (0.07)	1.28*** (0.08)	
	Car access. to Jobs	0.12*** (0.01)	0.19** (0.05)		-0.59*** (0.11)
	(Log) dist. to CBD	-0.70*** (0.02)	-0.39*** (0.09)	0.82*** (0.17)	
	% High inc. HH	-0.02*** (0.004)			
	% High edu. HH	-0.0008 (0.009)			0.18 (0.13)
	(Log) houses prices	0.0001*** (3.1e-05)		-1.9e-05 (0.0001)	-8.1e-06 (0.0001)
	Adj. LL Index	0.27	0.34	0.47	0.21
	AIC	13 020	1 215	925	1 406

Table E.5 – Employment location choice sub model (between brackets: standard deviation; significance level: * = $\alpha \leq 0.05$, ** = $\alpha \leq 0.01$, *** = $\alpha \leq 0.001$)

APPENDIX F

Appendices of Chapter 7

F.1 Additional tables and figures

Category	**Criteria**
Physical	Population density/threshold Built-up areas continuity Dwelling market
Socio-Economic	Employment threshold Jobs/Inhabitants ratio Median Income Primary sector share
Attractiveness and transport	Home-to-work commuting Home-to-school commuting
Dynamic	Population growth Built-up areas growth In/out migrations

Table F.1 – Criteria used for cities' delineations (adapted from Dujardin et al., 2007)

	Urban core	**Urban region**	**Extended urban area**
Main criteria	Morphological (built-up areas contiguity)	Socio-economic (population and jobs density), dynamic (population or jobs growth)	Attractivity (home-to-work commuting)
Area included	Narrow: densely built city centre	Medium-sized: city centre, suburbs and periurban areas	Large: all areas having functional relation with the CBD
Examples	Pole urbain (France), Morphological agglomeration (Belgium), Central counties (US)	Urban region (Belgium)	Anneau peri urbain (France), Metropolitan labour area (Belgium), Outlying counties (US)

Table F.2 – Typology of cities' functional delineations

Agents	**Variable**	**Parameter estimate**
Households	HH with car * car accessibility	Positive
	HH with univ. degree * % of univ. degree holder in the zone	Positive
	Green area score	Positive
	Low income HH * % of high income HH in the zone	Negative
	High income HH * % of high income HH in the zone	Positive
	HH with worker * (log of) diet. to CBD	Negative
	HH without car * rail station at less than 1 km	Positive
	Log of houses price	Negative
	Municipality of Brussels	Positive
Jobs	Log of (total) job density	Negative or Positive
	Log of job density (per activity sector)	Negative or Positive
	Log of non residential surface	Positive
	Population density	Negative
	Car accessibility	Positive
	% of high income HH	Positive

Table F.3 – Location choice sub models (Brussels case study, information derived from Cabrita et al., 2015; for jobs, the parameter estimates may vary from one activity sector to another)

Component	Variable	Availability Brussels	Paris	Zürich	Endogenous
Net income	Yearly wage before taxes	Yes	Yes	Yes	No
	Property income	No	No	No	No
	Local taxes on income	No	No	No	No
	Global taxes on income	No	Yes	No	No
	Local taxes on property value	No	Yes	No	No
	Global taxes on property value	No	No	No	No
Housing cost	Yearly rent/Property cost	Yes	Yes	Yes	Yes
	Maintenance cost	No	Yes	No	No
Travel cost	Total travel time	Yes	Yes	Yes	Yes
	Value of time	Yes	Yes	Yes	No
	Total travel cost	Yes	Yes	Yes	Yes
	Total travel distance	Yes	Yes	Yes	Yes
	Public transport cost	Yes	Yes	Yes	No
	Public transport comfort	No	No	No	No
Environment	Total travel distance by mode and fuel	No	No	No	No
	Floor space heated	Yes	Yes	Yes	Yes
	Type of energy	No	Yes	No	No
	Electricity' consumption	No	No	No	No
	Electricity generation	No	No	No	No

Table F.4 – Components of the social welfare (adapted from Proost et al., 2015)

Bibliography

23 US Code Art 134, 1962. Metropolitan transportation planning. https://www.law.cornell.edu/uscode/text/23/134 Accessed 9 nov. 2015.

ACP, 2000. North Livermore: Last Change for Smart Growth. Alameda County Planning. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1400/filename/1401.pdf.

ADT-ATO, 2015. Zones stratégiques. ADT-ATO. http://www.adt-ato.brussels/fr/zones-strat%C3%A9giques.

AFT, 1995. Alternatives for Future Urban Growth in California's Central Valley. American Farmland Trust. http://www.farmlandinfo.org/sites/default/files/FUTURE_URBAN_GROWTH_IN_CALIFORNIAS_CENTRAL_VALLEY_2ND_PRINTING_1.pdf.

Ahlfeldt, G., 2011. If Alonso was right: modeling accessibility and explaining the residential land gradient. Journal of Regional Sciences 51, 318–338.

Alamá-Sabater, L., Artal-Tur, A., Navarro-Azorín, J.M., 2011. Industrial location, spatial discrete choice models and the need to account for neighbourhood effects. The Annals of Regional Science 47, 393–418.

Allen, P.M., Sanglier, M., 1981a. A Dynamic Model of a Central Place System—II. Geographical Analysis 13, 149–164. doi:10.1111/j.1538-4632.1981.tb00722.x.

Allen, P.M., Sanglier, M., 1981b. Urban Evolution, Self-Organization, and Decisionmaking. Environment and Planning A 13, 167–183. doi:10.1068/a130167.

Alonso, W., 1964. Location and Land Use. Harvard University Press.

Amrhein, C.G., 1995. Searching for the Elusive Aggregation Effect: Evidence from Statistical Simulations. Environment and Planning A 27, 105–119. doi:10.1068/a270105.

Anas, A., 1982. Residential location markets and urban transportation: economic theory, econometrics, and policy analysis with discrete choice models. Studies in urban economics, Academic Press.

Anas, A., 2013. A summary of the applications to date of RELU-TRAN, a microeconomic urban computable general equilibrium model. Environment and Planning Part B 40, 959–970.

Anas, A., Arnott, R., Small, K., 1998. Urban spatial structure. Journal of Economic Litterature 36, 1426–1464.

Anas, A., Lindsey, R., 2011. Reducing Urban Road Transportation Externalities: Road Pricing in Theory and in Practice. Review of Environmental Economics and Policy 5, 66–88. doi:10.1093/reep/req019.

Anas, A., Liu, Y., 2007. A regional economy, land use and transportation model (RELU-TRAN): formulation algorithm design, and testing. Journal of Regional Science 47, 415–455.

Anderstig, C., Mattsson, L.G., 1992. Appraising large-scale investments in a metropolitan transportation system. Transportation 19, 267–283.

Anselin, L., 1988a. Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity. Geographical Analysis 20, 1–17. doi:10.1111/j.1538-4632.1988.tb00159.x.

Anselin, L., 1988b. Spatial econometrics: methods and models. Kluwer Academic Pub.

Anselin, L., 1989. What is special about spatial data?: alternative perspectives on spatial data analysis. National Center for Geographic Information and Analysis Santa Barbara, CA. http://www.odum.unc.edu/content/pdf/anselin%201989.pdf.

Anselin, L., 1995. Local indicators of spatial association - LISA. Geographical Analysis 27, 93–115.

Anselin, L., 2002. Under the hood: Issues in the specification and interpretation of spatial regression models. Agricultural Economics 27, 247–267.

Anselin, L., 2013. Spatial econometrics: methods and models. volume 4. Springer Science & Business Media.

Anselin, L., Lozano-Gracia, N., 2009. Spatial hedonic models. Palgrave handbook of econometrics 2, 1213–1250.

Anselin, L., Syabri, I., Kho, Y., 2006. GeoDa: An Introduction to Spatial Data Analysis. Geographical Analysis 38, 5–22. doi:10.1111/j.0016-7363.2005.00671.x.

Arauzo-Carod, J.M., Antolín-Manjón, M.C., 2004. Firm size and geographical aggregation: an empirical appraisal in industrial location. Small Business Economics 22, 299–312.

Arauzo-Carod, J.M., Liviano-Solis, D., Manjon, M., 2010. Empirical studies in industrial location: an assessment of their methods and results. Journal of regional sciences 50, 685–711.

Arbia, G., 1989. Statistical effects of spatial data transformations: a proposed general framework. Taylor and Francis.

Arbia, G., Petrarca, F., 2011. Effects of MAUP on spatial econometrics models. Letters in Spatial Ressources Sciences 4, 173–185.

Arbia, G., Petrarca, F., 2013. Effects of scale in spatial interaction models. Journals of Geographical Systems 15, 249–264.

Archer, W., Smith, M., 1993. Why do suburban office cluster? Geographical Analysis 25, 53–64.

Arrow, K., Dasgupta, P., Goulder, L., Daily, G., Elhrlich, P., Hea, G., Levin, S., Mailer, K.G., Schneider, S., Starret, D., Walker, B., 2004. Are we consuming too much? Journal of economic perspectives 18, 147–172.

Atkinson, G., Mourato, S., 2006. Cost-benefit analysis and the environment: recent developments. Technical Report. OCDE. http://www.lne.be/themas/beleid/milieueconomie/downloadbare-bestanden/ME11_cost-benefit%20analysis%20and%20the%20environment%20oeso.pdf.

Axhausen, K.W., Gärling, T., 1992. Activity?based approaches to travel analysis: conceptual frameworks, models, and research problems. Transport Reviews 12, 323–341. http://dx.doi.org/10.1080/01441649208716826.

Badoe, D., Miller, E., 2000. Transportation-land-use interaction: empirical findings in North America and their implications for modelling. Transportation Research Part D 5, 235–263.

de la Barra, T., 1989. Integrated land use and transport modelling. Cambridge University Press, Cambridge.

Barthelemy, J., Toint, P., 2013. Synthetic population generation without a sample. Transportation Science 47, 266–279. doi:10.1287/trsc.1120.0408.

Bartholomew, K., 2005. Integrating land use issues into transportation planning: scenario planning. Technical Report. University of Utah. http://faculty.arch.utah.edu/bartholomew/Bibliography.pdf.

Bartholomew, K., 2007. Land use-transportation scenario planning: promise and reality. Transportation 34, 397–412. doi:10.1007/s11116-006-9108-2.

Bates, J., Brewer, M., Hanson, P., Mcdonald, D., Simmonds, D., 1991. Building a strategic model for Edinburgh, in: Proceedings of seminar G: Transportation plannings methods, PTRC, Glenthorne House, London. pp. 161–185.

Batty, M., 1994. A chronicle of scientific planning: the anglo-american modelling experience. Journal of the American Planning Association 60, 7–16.

Batty, M., 2005. Cities and complexities: understanding cities with cellular automata, agent-based models, and fractals. The MIT Press, Cambridge.

Batty, M., 2008. Fifty Years of Urban Modeling: Macro-Statics to Micro-Dynamics, in: Albeverio, S., Andrey, D., Giordano, P., Vancheri, A. (Eds.), The Dynamics of Complex Urban Systems. Physica-Verlag HD, pp. 1–20. doi:10.1007/978-3-7908-1937-3_1.

Batty, M., 2014. Can it happen again? Planning support, Lee's Requiem and the rise of the smart cities movement. Environmnent and Planning B: Planning and Design 41, 388–391.

Batty, M., Longley, P., 1994. Fractal cities: a geometry of form and function. Academic press, San Diego and London.

Batty, M., Vargas, C., Smith, D., Serras, J., Reades, J., Johansson, A., 2013. SIMULACRA: fast land-use-transportation models for the rapid assessment of urban futures. Environment and Planning Part B 40, 987–1002.

Baudewyns, D., 1999. La localisation intra urbaine des firmes: une estimation logit multinomiale. Revue d'Économie Régionale et Urbaine 5, 915–930.

Baudewyns, D., Sekkat, K., Ben-Ayad, M., 2000. Infrastructure publique et localisation des entreprises a Bruxelles et en Wallonie, in: Beine, M., Docquier, F. (Eds.), Convergence des Régions: cas des régions belges. De Boeck, pp. 280–303.

Ben Akiva, M.E., Lerman, S.R., 1985. Discrete choice analysis: Theory and application to travel demand. MIT Press.

Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P., 2015. Integrated transport and land use modelling for sustainable cities. EPFL Press, Lausanne.

Bivand, R., Hauke, J., Kossowski, T., 2013. Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. Geographical Analysis 45, 150–179.

Bivand, R., Piras, G., 2015. Comparing Implementations of Estimation Methods for Spatial Econometrics. Journal of Statistical Software 63, 1–36.

Blondel, V., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large network. Journal of Statistical Mechanics: Theory and Experiment P10008.

Blondel, V., Krings, G., Thomas, I., 2010. Régions et frontières de téléphonie mobile en Belgique et dans l'aire métropolitaine bruxelloise. Brussels Studies 42, 1–13.

Boardman, A., Greenberg, D., Vining, A., Weimer, D., 2006. Cost-benefit analysis: concepts and practice. Prentice Hall, New Jersey.

Bode, E., 2008. Delineating Metropoltian Areas using Land Prices. Journal of Regional Sciences 48, 131–163.

Bonnel, P., Coulombel, N., Prados, E., Sturm, P., Arnaud, E., Boittin, C., Bouzouina, L., Cabrera Delgado, J., Capelle, T., Delons, J., Gilquin, L., Hilaire, V., De Lapparent, M., Nguyen-Luong, D., Nicolas, J.P., Saujot, M., Tschirhard, A., Vidard, A., 2014. A survey on the calibration and validation of integrated land use and transportation models, in: Symposium "Towards integrated modelling of urban systems", Lyon, France. https://hal.inria.fr/hal-01093254/file/survey_calibration_validation.pdf.

Boon, F., Gayda, S., 2000. Quelques impact du futur RER sur la Région Bruxelloise et sa Péripherie, in: 14ieme Congrès des économistes belges de langues francaise. http://www.stratec.be/sites/default/files/files/IMPACTSRER2-FB.pdf.

Borning, A., Waddell, P., Förster, R., 2008. Urbansim: Using Simulation to Inform Public Deliberation and Decision-Making, in: Chen, H., Brandt, L., Gregg, V., Traunmüller, R., Dawes, S., Hovy, E., Macintosh, A., Larson, C.A. (Eds.), Digital Government: E-Government Research, Case Studies, and Implementation. Springer US, Boston, MA, pp. 439–464. doi:10.1007/978-0-387-71611-4_22.

Bosredon, M., Doson, A.C., Simmonds, D., 2009. Transport/economic/land-use model of Scotland: land-use modeling with DELTA, in: Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management, Hong-Kong. http://www.davidsimmonds.com/files/2-Bosredon-CUPUM-2009-TELMoS.pdf.

Bowen, W.M., Mikelbank, B.A., Prestegaard, D.M., 2001. Theoretical and Empirical Considerations Regarding Space in Hedonic Housing Price Model Applications. Growth and Change 32, 466–490. doi:10.1111/0017-4815.00171.

Bowman, J.L., Ben-Akiva, M.E., 2000. Activity-based disaggregate travel demand model system with activity schedules. Transportation research part A 35, 1–28. http://www.civil.ist.utl.pt/~martinez/PDF/MobiCredit/Paper23.pdf.

Briant, A., Combes, P.P., Lafourcade, M., 2010. Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? Journal of Urban Economics 67, 287–302.

te Brömmelstroët, M., Pelzer, P., Geertman, S., 2014. Commentary. Forty years after Lee's requiem: are we beyond the seven sins. Environment and Planning B: Planning and Design 41, 381–387.

Brown, B.J., Hanson, M.E., Liverman, D.M., Merideth, R.W., 1987. Global sustainability: toward definition. Environmental management 11, 713–719.

BRTB, 2003. Vision 2030: Shaping Our Region's Future Together - Final Report. Baltimore Regional Transportation Board. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1418/filename/1419.pdf.

Brück, L., 2002. La périurbanisation en Belgique: comprendre le processus d'étalement urbain. Technical Report. SEGEFA-LMG. http://www.reseau-pwdr.be/media/3310/La%20p%C3%A9riurbanisation%20en%20Belgique%20comprendre%20le%20processus%20de%20l'%C3%A9talement%20urbain%20%281%C3%A8re%20partie%29.pdf.

Bulkeley, H., Betsill, M., 2005. Rethinking sustainable cities: multilevel governance and the "urban" politics of climate change. Environmental politics 14, 42–63.

Burgess, E.W., 1925. The growth of the city, in: Park, R.E., Burgess, E.W., McKenzie, R.D. (Eds.), The city. The university of Chicago press, Chicago and London, pp. 47–62.

Burgos, J., 1994. Integrated land-use and transport models in the Basque Country. Environment and Planning Part B 21, 603–610.

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimode inference: a practical information-theoretic approach. Springer, New-York.

CAAA, 1990. 1990 Clean Air Act Amendment Summary. http://www2.epa.gov/clean-air-act-overview/1990-clean-air-act-amendment-summary Accessed 9 nov. 2015.

Cabrita, I., Gayda, S., Hurtubia, R., Efthymiou, D., Thomas, I., Peeters, D., Jones, J., Cotteels, C., Nagel, K., Nicolai, T., Roder, D., 2015. Integrated land use and transport microsimulation for Brussels, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 373–412.

Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics - Theory and Methods 3, 1–27.

Caruso, G., Cavailhès, J., Peeters, D., Thomas, I., Frankhauser, P., Vuidel, G., 2015. Greener and larger neighbourhoods make cities more sustainable! A 2D urban economics perspective. Computers, Environment and Urban Systems 54, 82 – 94. doi:http://dx.doi.org/10.1016/j.compenvurbsys.2015.06.002.

Caruso, G., Peeters, D., Cavailhès, J., Rounsevell, M., 2007. Spatial configurations in a periurban city. A cellular automata-based microeconomic model. Regional Science and Urban Economics 37, 542–567.

Cavailhès, J., Peeters, D., Sekeris, E., Thisse, J.F., 2004. The periurbancity: why to live between the suburbs and the countryside. Regional Sciences and Urban Economics 34, 681–703.

Cavailhès, J., Thomas, I., 2013. Are Agricultural and Developable Land Prices Governed by the Same Spatial Rules? The Case of Belgium. Canadian Journal of Agricultural Economics/Revue canadienne d'agroéconomie 61, 439–463. doi:10.1111/j.1744-7976.2012.01265.x.

CBF, 1996. A Network of Livable Communities: Evaluating Travel Behavior Effect of Alternative Transportation and Community Designs for the National Capital Region. Chesapeake Bay Foundation. http://utah-primoprod.hosted.exlibrisgroup.com/primo_library/libweb/action/dlDisplay.do?vid=MWDL&afterPDS=true&docId=digcoll_uuu_11FHWA/1575.

CBRE, 2012. Brussels Office MarketView. CBRE. http://www.cbre.be/be_en/research/be_enResearch%20%20Content/BrusselsOfficeMarketView_Q42012.pdf.

CCC, 2003. Shaping Our Future. Contra Costa County. http://content.lib.utah.edu/cdm/ref/collection/FHWA/id/1380.

CCMPO, 2004. 2025 Chittenden County Metropolitan Transportation Plan. Chittenden County Metropolitan Planning Organization. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1419/filename/1420.pdf.

CCoD, 2002. Blueprint Denver: An Integrated Land Use and Transport Plan. City and County of Denver. https://www.denvergov.org/Portals/646/documents/planning/blueprintdenver/BlueprintDenver.pdf.

CEE, 1999. Two Roads Diverge - Final Report. Center for Energy and Environment. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1413/filename/1414.pdf.

Chang, K.T., Khatib, Z., Ou, Y., 2002. Effects of zoning structure and network detail on traffic demand modeling. Environment and Planning B: Planning and Design 29, 37–52.

Charlton, M., Fotheringham, A.S., 2009. Geographically weighted regression white paper. Technical Report. National Centre for Geocomputation, National University of Ireland, maynooth. http://www.geos.ed.ac.uk/~gisteac/fspat/gwr/gwr_arcgis/GWR_WhitePaper.pdf.

Cheshire, P.C., 2010. Why Brussels need a city region for the city. Technical Report. E-Books 7, Re-Bel Initiative.

Cheshire, P.C., Gornostaeve, G., 2002. Villes et régions urbaines: des comparaisons fiables doivent reposer sur des territoires comparables. Les cahiers de l'IAURIF 135, 13–32.

Clark, C., 1951. Urban population densities. Journal of the royal statistics society. Series A (general) 114, 490—496.

CoA, 2000. Planned Growth Strategy Findings Report. City of Albuquerque. http://www.cabq.gov/council/documents/pgs/Part1.pdf.

Coombes, M., 2013. From City-region Concept to Boundaries for Governance: The English Case. Urban Studies doi:10.1177/0042098013493482.

Coombes, M., Bond, S., 2007. Travel-to-work areas: the 2007 review. Technical Report. Office for national statistics. London. www.istat.it/it/files/2014/12/final_TTWA_report.doc.

CoP, 1997. North Land Use Plan. City of phoenix. https://www.phoenix.gov/pddsite/Documents/pdd_pz_pdf_00087.pdf.

Coppola, P., Ibeas, A., dell'Olio, L., Cordera, R., 2013. LUTI model for the metropolitan area of Santander. Journal of Urban Planning and Development 139, 153–165.

Cornelis, E., Barthelemy, J., Paully, X., Walle, F., 2012. Modélision de la mobilité résidentielle en vue d'une micro-simulation des evolutions de population. Les cahiers scientifiques du transport 62, 65–84.

Cörvers, F., Hensen, M., Bongaerts, D., 2009. Delimitation and coherence of functional and administrative regions. Regional Studies 43, 19–31.

Country, C., 2003. Flagstaff Area Regional Land Use and Transportation Plan. Coconino County. http://www.flagstaff.az.gov/DocumentCenter/View/366.

CPF, 2002. Highway 41 Corridor Master Plan: A Guide for Land Use and Transportation Improvements. City of Post Falls. http://www.postfallsidaho.org/PZDept/pzforms/Planning/Hwy41CorridorMaster.pdf.

CRCG, 2003. Integrated Infrastructure Planning Project. Catawba Regional Council of Governements. http://content.lib.utah.edu/cdm/ref/collection/FHWA/id/1410.

CRI, 2002. Room for Growth, Room for Open Space: Planning for a Sustainable Future. Conservation Research Institute. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1384/filename/1385.pdf.

Croissant, Y., 2012. mlogit: multinomial logit model. R package version 0.2-3.

CUF, 1999. QGET, Scenario Analysis: Executive Summary. Coalition for Utah's Future. http://www.epa.gov/smartgrowth/pdf/envision_utah.pdf.

CUF, 2004. Envision Utah, Mountain View Corridor Growth Choices Process: Helping Solve Our Communities' Transportation Problems. Coalition for Utah's Future. .

Delloye, J., Peeters, D., Thomas, I., 2015. On the Morphology of a Growing City: A Heuristic Experiment Merging Static Economics with Dynamic Geography. PLoS ONE 10, e0135871. doi:10.1371/journal.pone.0135871.

Dembour, C., 2004. Quelles frontières pour l'agglomération Bruxelloise. Technical Report. CEREC Working Paper. http://centres.fusl.ac.be/CEREC/document/2004/cerec2004_6.pdf.

DGSIE, 2011. NACE-Bel: Nomenclature d'activité. http://statbel.fgov.be/fr/binaries/FR%20Nace%202008%20avec%20notes%20explicatives_tcm326-65642.pdf. Accessed: 2015-11-04.

DGSIE, 2015a. Census 2011 Belgique. http://census2011.fgov.be/index_fr.html.

DGSIE, 2015b. Prix de l'immobilier. http://economie.fgov.be/fr/statistiques/chiffres/economie/construction_industrie/immo/.

Di Pasquale, D., Wheaton, W.C., 1996. Urban economics and real estate markets. Prenctice Hall, New Jersey.

Di Zio, S., Montanari, A., Staniscia, B., 2010. Simulation of urban development in the City of Rome: Framework methodology and problem solving. Journal of Transport and Land Use 3, 85–105.

de Dios OrtÃozar, J., Willumsen, L.G., 2011. Modelling transport. John Wiley & Sons.

Dowling, R.G., 2005. Predicting air quality effects of traffic-flow improvements: final report and user's guide. volume 535. Transportation Research Board National Research. http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_rpt_535.pdf.

DRCG, 1995. Metro Vision 2020 Urban Form Alternative and Evaluation Criteria. Denver Regional Council of Governments. https://drcog.org/documents/2020_Metro_Vision_Plan-1.pdf.

Duda, R.O., Hart, P.E., 1973. Pattern classification and scene analysis. Jon Wiley and Sons.

Dujardin, C., 2001. Effet de frontière et intégration spatiale: les migrations alternantes et la frontière linguistique en Belgique. L'espace géographique 2001, 307–320.

Dujardin, C., Selod, H., Thomas, I., 2008. Residential Segregation and Unemployment: The Case of Brussels. Urban Studies 45, 89–113.

Dujardin, C., Thomas, I., Tulkens, H., 2007. Quelles frontières pour Bruxelles? Une mise a jour. Reflets et Perspectives de la Vie Economique 46, 155–176.

DVRPC, 2002. Planned Projects Status Report. Delaware Valley Regional Planning Commission. http://www.dvrpc.org/reports/08093.pdf.

DVRPC, 2003. Regional Analysis of What-If Transportation Scenarios - Final Report. Delaware Valley Regional Planning Commission. http://www.dvrpc.org/reports/03020.pdf.

EC, 2015. Council buildings. Council of the European Union. http://www.consilium.europa.eu/en/contact/address/council-buildings/.

Echenique, M., 2001. Transport and land-use interaction. Part A: integrated modelling methology. Technical Report. Report EUR20124 EN.

Echenique, M., Flowerdew, A., Hunt, J., Mayo, T., Skidmore, I., Simmonds, D., 1990. The MEPLAN model of Bilbao, Leeds and Dortmund. Transport Reviews 10, 309–322.

ECT, 2003. Envision Central Texas Briefing Packet. Envision Central Texas. http://content.lib.utah.edu/cdm/ref/collection/FHWA/id/1421.

EEA, 2006. CORINE Land Cover. http://www.eea.europa.eu/publications/COR0-landcover. Accessed: 2015-11-02.

EEA, 2013. Urban environment. European Environmental Agency. http://www.eea.europa.eu/themes/urban/intro.

Efthymiou, D., Antoniou, C., 2013. How do transport infrastructure and policies affect houses price and rents? Evidence from Athens, Greece. Transportation Research Part A 52, 1–22.

Efthymiou, D., Farooq, B., Bierlaire, M., Antoniou, C., 2014. Multidimensional indicators analysis for transportation policy evaluations. Transport research record, journal of the transportation research board , 83–94.

Elhorst, J.P., 2010. Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis 5, 9–28. doi:10.1080/17421770903541772.

Eliasson, J., 2009. A cost-benefit analysis of the Stockholm congestion charging system. Transportation Research Part A: Policy and Practice 43, 468–480.

EPA, 2000. Environmental Assessment: 17th Street Extension and Atlantic Steel Redevelopment Project. EPA. http://content.lib.utah.edu/cdm/ref/collection/FHWA/id/1416.

Eradus, P., Shoemakers, A., van der Hoorn, T., 2002. Four applications of the TIGRIS model to the Nertherlands. Journal of Transport Geography 10, 111–121.

Ettema, D., 2011. A multi-agent model of urban processes: Modelling relocation processes and price setting in housing markets. Computers, Environment and Urban Systems 35, 1 – 11. doi:http://dx.doi.org/10.1016/j.compenvurbsys.2010.06.005.

Farmer, C., Fotheringham, A.S., 2011. Network-based functional regions. Environment and Planning A 43, 2723–2741.

Farooq, B., Hurtubia, R., Bierlaire, M., 2015a. Simulation based generation of a synthetic population for Brussels, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 95–112.

Farooq, B., Muller, K., Bierlaire, M., Axhausen, K.W., 2015b. Methodologies for synthesising populations, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 77–94.

Federal Register, 2000. Standards for defining metropolitan and micropolitan statistical areas. Federal Register 65.

Federal Register, 2002. Qualifying urban areas for Census 2000: Notices. Federal Register 67.

Feldman, O., Simmonds, D., Zachariadis, V., Richmond, E., Mackett, R., 2007. SIMDELTA - A micro simulation approach to household location modelling, in: Proceedings of the World Conference on Transport Research, University of California, Berkeley. http://www.davidsimmonds.com/files/resourcesmodule/@random4767836fd9d18/1199790463_Microsim_revised.pdf.

Fischer, M.M., Getis, A., 2010. Handbook of applied spatial analysis: software tools, methods and applications. Springer.

FMCoG, 1998. Fargo-Moorhead Area Short and Long Range Metro Transportation Plan. Fargo-Moorhead Council of Governments. http://www.fmmetrocog.org/new/assets/metro2040/.

Fotheringham, A., Brunsdon, C., Charlton, M., 2000a. Quantitative geography: perspectives on spatial data analysis. Sage Publications Ltd.

Fotheringham, A.S., Brundson, C., Charlton, M., 2002. Geographically weighted regression. Chicester, Wiley.

Fotheringham, A.S., Brunsdon, C., Charlton, M., 2000b. Quantitative geography: perspectives on spatial data analysis. SAGE Publications Ltd.

Fotheringham, A.S., Rogerson, A., 2009. The SAGE handbook of spatial analysis. SAGE Publications limited, London, UK.

Fotheringham, A.S., Wong, D.W., 1991. The modifiable areal unit problem in multivariate statistical analysis. Environment and planning A 23, 1025–1044.

Fujita, M., 1989. Urban Economic Theory. Cambridge university press.

Fujita, M., Ogawa, H., 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. Regional sciences and urban economics 12, 161–196.

Gallay, O., 2010. Starting with UrbanSim: on the creation of an introductory project. Technical Report. Transp-OR, EPFL. http://infoscience.epfl.ch/record/152471/files/Gallay_UrbanSim_2010.pdf.

Gehlke, C., Biehl, K., 1934. Certain effects of grouping upon the size of the correlation coefficient in census tract material. Journal of the American Statistical Association 29, 169–170.

Gerber, P., 2012. Advancement in conceptualizing cross-border daily mobility: the Benelux context in the European Union. European Journal of Transport and Infrastructure Research 12, 178–197.

Gerber, P., Klein, O., Carpentier, S., 2012. Local Mobility and Cross-border Peri-urbanisation, in: Sohn, C. (Ed.), An Emerging Cross-border Metropolitan Region. Peter Lang, Brussels, pp. 141–159.

Getis, A., Aldstadt, J., 2004. Constructing the Spatial Weights Matrix Using a Local Statistic. Geographical Analysis 36, 90–104. doi:10.1111/j.1538-4632.2004.tb01127.x.

Geurs, K.T., van Wee, B., 2004. Land-use/Transport interactions models as tools for sustainability impact assessment of transport investments: review and research perspectives. European journal of transport infrastructure research 4, 333–355.

Gibbons, S., Overman, H.G., 2010. Mostly Pointless Spatial Econometrics? SERC Discussion Papers 0061. Spatial Economics Research Centre, LSE.

Giuliano, G., Small, K., 1991. Subcenters in the Los Angeles Region. Regional Sciences and Urban Economics 21, 163–182.

Glaeser, E.L., Kahn, M.E., 2003. Sprawl and Urban Growth. Working Paper 9733. National Bureau of Economic Research. doi:10.3386/w9733.

Glaeser, E.L., Kahn, M.E., 2008. The Greenness of Cities: Carbon Dioxide Emissions and Urban Development. Working Paper 14238. National Bureau of Economic Research. doi:10.3386/w14238.

Goffette-Nagot, F., Reginster, I., Thomas, I., 2011. Spatial analysis of residential land prices in Belgium: accessibility, linguistic border, and environmental amenities. Regional Studies 45(9), 1253–1268.

Goldner, W., 1971. The Lowry Model Heritage. Journal of the American Institute of Planners 37, 100–110.

Goodchild, M., Gopal, S., 1989. The accuracy of spatial databases. Taylor and Francis, UK.

Goodchild, M., Marvin, S., 1996. GIS and environmental modelling: progress and issues. Oxford University Press, New York.

Goodchild, M., Parks, B., Steyaert, T., 1993. Environmental modelling with GIS. Oxford University Press, New York.

GRTA, 2003. Northern Sub-Area GA 400. Georgia Regional Transportation Authority. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1417/filename/1418.pdf.

GT, 1998. Regional Development Choices Project. Greater Triangle. http://www.tjcog.dst.nc.us/rdc2.htm.

Guimaraes, P., Figueiredo, O., Woodward, D., 2004. Industrial location modelling: extending the random utility framework. Journal of regional science 44, 1–20.

Guo, J.Y., Bhat, C.R., 2004. Modifiable areal units: Problem or perception in modeling of residential location choice? Transportation Research Record: Journal of the Transportation Research Board 1898, 138–147.

Guo, J.Y., Bhat, C.R., 2007. Operationalizing the concept of neighborhood: application to residential location choice analysis. Journal of Transport Geography 15, 31–45.

Guzman, L., de la Hoz, D., Circella, G., 2014. Evaluations of synergies from transportation policy packages using a social welfare maximization approach: a case study for Madrid, Spain. Case Studies on Transport Policy In press.

Haining, R., 1993. Spatial data analysis in the social and environmental sciences. Cambridge University Press.

Handy, S., 2008. Regional transportation planning in the US: an examination of changes in technical aspects of the planning process in response to changing goals. Transport policy 15, 113–126.

Hansen, W., 1959. How accessibility shapes land use. Journal of the American Institute of Planners 25, 73–76.

Harris, B., 1965. New tools for planning. Journal of the American Institute of Planners 31, 90–95.

Harris, B., 1985. Urban simulation models in regional science. Journal of Regional Science 25, 545–567.

Harris, B., 1994. The real issues concerning Lee's requiem. Journal of the American Planning Association 60, 31–34.

Harris, C.D., Ullman, E.L., 1945. The Nature of Cities. The Annals of the American Academy of Political and Social Science 242, 7–17.

Hayter, R., 1997. The dynamic of industrial location. The factory, the firm, and the production system. Wiley.

Hély, V., Antoni, J.P., 2014. A theoretical approach to combine sustainability indicators for decision making in planning issues, in: 12^{th} International conferences on design decision systems in architecture and urban planning, Eindhoven, The Netherlands.

Heppenstall, A., Smith, D., 2014. Spatial Microsimulation, in: Fischer, M.M., Nijkamp, P. (Eds.), Handbook of Regional Science. Springer-Verlag, Berlin, pp. 1235–1252.

Hoyt, H., 1939. The structure and growth of residential neighborhoods in american cities. Federal housing administration.

Hoyt, H., 1965. Recent distortions of the classical models of urban structure. Land economics 40, 199—212. `http://www.jstor.org/stable/pdf/3144351.pdf?acceptTC=true`.

HRPDC, 2003. Hampton Roads Smart Growth Analysis. Hampton Roads Planning District Commission. `http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1425/filename/1426.pdf`.

Hull, A., Alexander, E.R., Khakee, A., Woltjer, J., 2011. Evaluation for participation and sustainability in planning. Routledge London ; New York.

Hunt, J., 1994. Calibrating the Naples land-use and transport model. Environment and Planning Part B 21, 569–590.

Hunt, J., Abraham, J., De Silva, D., 2009a. PECAS Software user guide. `http://www.hbaspecto.com/pecas/downloads`.

Hunt, J., Abraham, J., De Silva, D., 2009b. PECAS Theoretical formulation. `http://www.hbaspecto.com/pecas/downloads`.

Hunt, J., Abraham, J., De Silva, D., Zhong, M., Bridges, J., Mysko, J., 2008. Developing and applying a parcel-level simulation of developers action in Baltimore, in: Proceedings of the 87th annual meeting of the Transportation Research Board, Washington DC. `http://v8nu74s71s31g374r7ssn017uloss3c1vr3s.unbf.ca/~ming/document/Baltimore.pdf`.

Hunt, J., Echenique, M., 1993. Experiences in the application of the MEPLAN framework for land use and transport interaction modelling, in: Proceedings of the 4th National Conference on the Application of Transportation Planning Methods, Daytona Beach. pp. 723–754.

Hunt, J., Miller, E., Kriger, D., 2005. Current operational urban land-use transport modeling frameworks. Transport reviews 25, 329–376.

Isard, W., 1951. Interregional and Regional Input-Output Analysis: A Model of a Space-Economy. The Review of Economics and Statistics 33, 318–328.

Ishizaka, A., Nemry, P., 2013. Multi-criteria decision analysis: methods and software. John Wiley and Sons Ltd, New Jersey.

ISTEA, 1991. Intermodal Surface Transport Efficiency Act of 1991. http://ntl.bts.gov/DOCS/istea.html Accessed 9 nov. 2015.

Jantz, C.A., Goetz, S.J., 2005. Analysis of scale dependencies in urban land-use-change model. International Journal of Geographical Information Science 19(2), 219–241.

Jenerette, D.G., Wu, J., 2001. Analysis and simulation of land-use chang in the central Arizona - Phoenix region, USA. Landscape Ecology 16, 611–626.

Julien, P., 2000. Mesurer un univers urbain en expansion. INSEE premiere 336, 3–33.

Julien, P., 2001. Les grandes villes françaises étendent leur influence. INSEE premiere 766.

Kahn, M.E., 2000. The environmental impact of suburbanization. Journal of Policy Analysis and Management 19, 569–586. doi:10.1002/1520-6688(200023)19:4<569::AID-PAM3>3.0.CO;2-P.

Karlsson, C., Olsson, M., 2006. The identification of functional regions: theory, methods, and applications. The Annals of Regional Science 40, 1–18. doi:10.1007/s00168-005-0019-5.

Kesteloot, C., Turck, A., Vandermotten, C., 2001. Structure sociales et quartiers en difficultés dans les régions urbaines belges. Technical Report. Rapport au ministre délegué a la politique des grandes villes.

KF, 2012. Brussels residential report - new built. Knight Frank. http://content.knightfrank.com/research/464/documents/en/2012-1018.pdf.

Klapka, P., Halás, M., Erlebach, M., Tonev, P., Bednar, M., 2014. A multistage agglomerative approach for defining functional region of the Czech republic: the use of 2001 commuting data. Moravian geographical report 22, 2–13.

Kocabas, V., Dragicevic, S., 2006. Assessing cellular automata model behaviour using a sensitivity analysis approach. Computer, Environment and Urban Systems 30, 921–953.

Konjar, M., Lisec, A., Drobne, S., 2010. Methods for delineation of functional regions using data on commuters, in: 13^{th} AGILE International conference on Geographic Information Science, Guimaraes, Portugal.

Kostov, P., 2013. Choosing the right spatial weighting matrix in quantile regression model. ISRN Economics 2013. doi:10.1155/2013/158240. article ID 158240.

Kwan, M.P., 2012. The uncertain geographic context problem. Annals of the Association of American Geographers 102, 958–968.

Ladd, H., Wheaton, W., 1991. Causes and consequences of the changing urban form. Regional Sciences and Urban Economics 21, 157–162.

Landis, J., 1994a. The california urban futures model: a new generation of metropolitan simulation models. Environment and Planning B: Planning and Design 21, 399–420.

Landis, J., 1994b. Pilot study of solano and sonoma land use and development policy alternatives. Technical Report. Institute of urban and regional development. https://escholarship.org/uc/item/98c0043t.

Landis, J., 1995. Imagining Land Use Futures: Applying the California Urban Futures Model. Journal of the American Planning Association 61, 438–457.

Landis, J., Zhang, M., 1998a. The second generation of the california urban future model. Part I: model logic and theory. Environment and Planning B: Planning and Design 25, 657–666.

Landis, J., Zhang, M., 1998b. The second generation of the california urban future model. Part II: specification and calibration results of the land use change sub model. Environment and Planning B: Planning and Design 25, 795–824.

Le Gallo, J., 2002. Econométrie spatiale: l'autocorrélation spatiale dans les modèles de régression linéaire. Economie & prévision 155, 139–157.

Le Gleau, J.P., Pumain, D., Saint-Julien, T., 1996. Villes d'Europe: à chaque pays sa définition. INSEE Economie et statistiques 294–295, 9–23. =http://www.epsilon.insee.fr:80/jspui/handle/1/21370.

Le Jeannic, T., Vidalenc, J., 1997. Poles urbains et periurbnisation. Le zonage en aires urbaines. INSEE premiere 516.

Lee, D., 1973. Requiem for Large-Scale Models. Journal of the American Institute of Planners 39, 163–178.

Lee, D., 1994. Retrospective on large-scale urban models. Journal of the American Planning Association 60, 35–40.

Lee, D., 2009. 2009 TMA/MPO Modelling Activity survey. http://www.urbansim.org/pub/Main/WebHome/2009_MPO_Modeling_Survey_RESULT_TMIP.pdf.

Lemoy, R., Raux, C., Jensen, P., 2010. An agent-based model of residential patterns and social structure in urban areas. Cybergeo : European Journal of Geography doi:10.4000/cybergeo.23381. http://cybergeo.revues.org/23381.

LeSage, J.P., 1999. The theory and practice of spatial econometrics. University of Toledo. Toledo, Ohio 28, 33.

LeSage, J.P., Pace, R.K., 2009. Introduction to Spatial Econometrics. Chapman & Hall/CRC, Boca Raton, Florida.

Löchl, M., Axhausen, K.W., 2010. Modelling hedonic residential rents for land use and transport simulation while considering spatial effects. Journal of Transport and Land Use 3.

Lord, S., Gerber, P., 2013. Immigration, dynamiques socio-économiques territoriales et mouvements résidentiels. Quelles perspectives pour les résidents du Luxembourg ? Annales de géographie 690, 175–199. www.cairn.info/revue-annales-de-geographie-2013-2-page-175.htm.

Lowry, I., 1964. A Model of Metropolis. Technical Report. The RAND Corporation. http://www.prgs.edu/content/dam/rand/pubs/research_memoranda/2006/RM4035.pdf.

Lucotte, M., Nguyen, S., 2013. Equilibrium and advanced transportation modelling. Springer Science & Business Media.

Mackett, R., 1990. The systematic application of LILT model to dortmund, Leeds and Tokyo. Transport Reviews 10, 323–338.

MAG, 2002. MAG Regional Transportation Plan: Mobility for the new Millenium. Mariscopa Association of Governments. https://transportationforcommunities.com/cases/pdf/.

MARC, 2001. Smart Choices: Understanding the Cost of Development - An Element of Creating Quality Places. Mid-America Regional Council. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1394/filename/1395.pdf.

Marissal, P., Medina Lockhart, P., Vandermotten, C., Van Hamme, G., 2006. Les Structure Socio-Economiques de l'Espace Belge. Technical Report. Monographie de l'enquête socio-économique générale 2001.

Marschak, J., 1959. Binary Choice Constraints on Random Utility Indicators. Technical Report. Cowles Foundation for Research in Economics, Yale University.

Martinez, F., 2003. A stochastic land use equilibrium model, in: Moving through nets:The physical and social dimensions of travel, 10th International Conference on Travel Behaviour Research, Lucerne.

Martinez, F., Donoso, P., 2004. MUSSA: a behavioural land use ewuilibirum model with location externalities, planning regulations and pricing policies. http://www.cec.uchile.cl/~dicidet/fmartinez/Mussa.PDF.

Martinez, F., Donoso, P., 2010. The MUSSA II land use auction equilibrium model, in: Pagliara, F., Preston, J., Simmonds, D. (Eds.), Residential Location Choice. Models and Applications. Springer, Berlin.

May, A., Shepherd, S., Emberger, G., Ash, A., Zhang, X., Paulley, N., 2005. Optimal land use transport strategies: methodology and application to european cities. Transportation Research Record 1924, 129–138.

McCann, P., Sheppard, S., 2003. The rise, fall and rise again of industrial location theory. Regional studies 37, 649–663.

McFadden, D., 1978. Modelling the choice of residential location, in: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), Spatial Interaction Theory and Planning Models. North Holland, pp. 75–96.

McMillen, D.P., 1996. One hundred fifty years of land values in Chicago: a non parametric approach. Journal of Urban Economics 40, 100–124.

McMillen, D.P., 2001. Nonparametric employment subcenters identification. Journal of Urban Economics 50, 448–473.

McMillen, D.P., 2003. Spatial autocorrelation or model misspecification? International Regional Science Review 26, 208–217.

McMillen, D.P., Smith, S., 2003. The number of subcenters in large urban areas. Journal of urban economics 53, 321–338.

MCPB, 2002. Transportation Policy Task Force: Final Draft Report. Montgomery county planning board. http://www.montgomeryplanning.org/community/plan_areas/georgia_avenue/master_plans/upper_rock_creek/master_plan_approved_july_2004/documents/transportation_plan.pdf.

Meester, W., Pellenberg, P., 2006. The spatial preference map of Dutch entrepreneurs: subjective rating of locations, 1983, 1991, and 2003. Tijdschrijft for economische en sociale geografie 97(4), 364–376.

Menard, A., Marceau, D.J., 2005. Exploration of spatial scale sensitivity in geographic cellular automata model. Environment and Planning B: Planning and Design 32, 693–714.

METRO, 2004. Community Connections: A Transportation Vision for the Next 25 Years. METROPLAN Orlando. http://www.metroplanorlando.com/files/view/tr7-transit-system-concept-plan-final-adopted.pdf.

Miller, E., Kriger, D., Hunt, J., 1998. Integrated urban models for simulation of transit and land use policies. Technical Report. Transport Cooperative Research Project. http://www.nap.edu/read/9435/chapter/1.

Miller, E., Kriger, D., Hunt, J., 1999. Integrated urban models for simulation of transit and land use policies: guidelines for implementation and use. Technical Report. Transportation Research Board. http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_48.pdf.

Mills, E.S., 1967. An aggregative model of resource allocation in a metropolitan area. The American Economic Review , 197–210http://www.jstor.org/stable/pdf/1821621.pdf.

Moeckel, R., 2007. Bussiness decision locations and urban sprawl: a microsimulation of business relocation and firmography. Technical Report. Institute spatial planning.

Moniteur Belge, 2004. Ordonnance portant assentiment à la Convention du 4 avril 2003 entre l'Etat fédéral, la Région flamande, la Région wallone et la Région de Bruxelles-Capitale, visant à mettre en oeuvre le programme du réseau express r'egional de, vers, dans et autour de Bruxelles. Moniteur Belge.

MORPC, 2004. Regional Fact Book. Mid-Ohio Regional Planning Commission. http://www.morpc.org/pdf/RegionalConnectionFinalReport.pdf.

Munz, M., Weidlich, W., 1990. Settlement formation. The Annals of Regional Science 24, 177–196. doi:10.1007/BF01581848.

Muth, R.F., 1969. Cities and housing: the spatial pattern of urban residential land use.

Nagel, K., Grether, D., Beuck, U., Chen, Y., Riser, M., Axhausen, K., 2008. Multi-agent transport simulations and economic evaluations. Journal of economics and statistics 228, 173–194.

NATO, 2014. New NATO headquarters. North Atlantic Treaty Organisation. http://www.nato.int/cps/en/natohq/topics_49287.htm.

Newling, B.E., 1969. The spatial variation of urban population densities. Geographical review 59, 242—252.

Nguyen-Luong, D., 2008. An integrated land use-transport model for the Paris region (SIMAURIF): Ten lessons learned after four years of development. Technical Report. Institut d'Aménagement et d'Urbanisme de la Région Ile-de-France. Paris. http://www.iau-idf.fr/ultisim/IMG/pdf/Paper_SIMAURIF_10_lessons_TRB_2009.pdf.

Nicolai, T., Nagel, K., 2010. Coupling MATSim and UrbanSim: Software design issues. Technical Report. SustainCity Working Paper. http://www.sustaincity.org/publications.

Nicolai, T., Nagel, K., 2011. Invetigating accessibility indicators for feedback from a travel to a land use model. Technical Report. SustainCity Working Paper. http://www.sustaincity.org/publications.

Nicolai, T., Nagel, K., 2015. Integration of agent-based transport and land use models, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 333–354.

Nicolai, T., Wang, L., Nagel, K., Waddell, P., 2010. Coupling an urban simulation model with a travel mode - a first sensitivity test. Technical Report. SustainCity Working Paper. http://www.sustaincity.org/publications.

Nijkamp, P., Blaas, E., 1994. Impact assessment and evaluation in transportation planning. Kluwer academic publisher, Doordrecht, The Netherlands.

NJOSP, 1992. Impact Assessment of the New Jersey Interim State Development and Transportation Plan. New Jersey Office of State Planning. http://www.state.nj.us/state/planning/publications/084-impact-assessment-nj-state-plan-execsum-022892.pdf.

NJOSP, 2000. The Impact Assessment of the New Jersey State Plan. New Jersey Office of State Planning. http://www.nj.gov/state/planning/docs/iaexecsumm090100.pdf.

ODT, 2001. Modeling Analysis of Willamette Valley Transportation and Land Use Alternatives. Oregon Department of Transportation. http://www.oregon.gov/ODOT/TD/TP/docs/statewide/smap.pdf.

ONSS, 2015. Donnés statistiques concernant l'emploi par lieu de travail - complément. http://www.onssrszlss.fgov.be/fr/statistiques/statistiques-en-ligne. Accessed: 2015-11-02.

Openshaw, S., 1984. Ecological Fallacies and the Analysis of Areal Census Data. Environment and Planning A 16, 17–31. doi:10.1068/a160017.

Openshaw, S., Taylor, P.J., 1979. A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem, in: Wrigley, N. (Ed.), Statistical Applications in the Social Sciences. Plon, pp. 127–144.

Orcutt, G., 1960. Simulation of economic systems. The American Economic Review , 894–907http://www.jstor.org/stable/pdf/1810951.pdf.

Orcutt, G., Greenberger, M., Rivlin, A., Korbel, J., 1961. Microanalysis of socio-economic systems: a simulation study. Harper and Row, New-York.

ORNL, 2002. Land Use Technical Report. Oak Ridge National Laboratory. http://landuseplanning.ornl.gov/technical_report/UT-B_Final_Land_Use_Report.pdf.

Pagliara, F., Wilson, A., 2010. The State-of-the-Art in building Residential Location Models. Residential Location Choice 1, 1–20.

de Palma, A., Motamedi, K., Nguyen-Luong, D., Ouaras, H., Picard, N., 2008. SIMAURIF: the whole calibration and application to the north tangential railways project. Technical Report. SIMAURIF. http://www.ekf.vsb.cz/export/sites/ekf/projekty/cs/weby/esf-0116/databaze-prispevku/clanky_ERSA_2008/135.pdf.

de Palma, A., Motamedi, K., Picard, N., 2007. Accessibility and environmental quality: inequality in the Paris housing market. European Transport/Trasporti Europei 36, 47–74.

de Palma, A., Picard, N., Motamedi, K., 2015a. Application of UrbanSim in Paris (Ile-de-France) case study, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 413–460.

de Palma, A., Saifuzzaman, M., Motamedi, K., 2015b. Integration of dynamic transport models and agent-based land use models, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 355–369.

Patterson, Z., Bierlaire, M., 2010. Development of Prototypes UrbanSim models. Environment and Planning Part B 37, 344–366.

Patterson, Z., Kryvobokov, M., Marchal, F., Bierlaire, M., 2010. Disaggregate models with aggregate data: Two UrbanSim applications. Journal of Transport and Land Use 3.

Paulet, J.P., 2009. Géographie Urbaine. Armand Colin, Paris.

PDRCC, 2003. Planning Implications of Alternate Development Patterns on Infrastructure and Existing Planning Policies. Pee Dee Regional Council of Governments. `http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1411/filename/1412.pdf`.

Pfaffenbichler, P., Emberger, G., Sheperd, S., 2010. A system dynamics approach to land use-transport interaction modelling: the strategic model MARS and its application. System Dynamics Review 26, 262–282.

Pholo Bala, A., Peeters, D., Thomas, I., 2014. Spatial issues on a hedonic estimation of rents in Brussels. Journal of Housing Economics 25, 104 – 123. doi:`http://dx.doi.org/10.1016/j.jhe.2014.05.002`.

Proost, S., Van der Loo, S., Antoniou, C., Efthymiou, D., 2015. Indicators of sustainable development for microsimulation models, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne.

Pumain, C., 2003. Scaling laws and urban systems. MIMEO.

Rasouli, S., Timmermans, H., 2013. What-Ifs, If-Whats and Maybes: Sketch of Ubiquitous Collaborative Decision Support Technology, in: Geertman, S., Toppen, F., Stillwell, J. (Eds.), Planning Support Systems for Sustainable Urban Development. Springer Berlin Heidelberg. volume 195 of *Lecture Notes in Geoinformation and Cartography*, pp. 19–29. doi:`10.1007/978-3-642-37533-0_2`.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., Strogatz, S.H., 2010. Redrawing the Map of Great Britain from a Network of Human Interactions. PLoS ONE 5, e14248. doi:`10.1371/journal.pone.0014248`.

Redfearn, C., 2007. The topography of metropolitan employment: identifying centers of employment in a polycentric urban area. Journal of urban economics 61, 519–541.

Rémy, A., 2004. Morphologie urbane. Géographie, Améngagement et architecture de la ville. Armand Colin, Paris.

Rey, S., Anselin, L., 2007. PySAL: A Python library of spatial analytical methods. Review of Regional Studies 37, 5–27.

Reynolds, H., Amrhein, C., 1998. Some Effects of Spatial Aggregation on Multivariate Regression Parameters, in: Griffith, D., Amrhein, C., Huriot, J.M. (Eds.), Econometric Advances in Spatial Modelling and Methodology. Springer US. volume 35 of *Advanced Studies in Theoretical and Applied Econometrics*, pp. 85–106. doi:10.1007/978-1-4757-2899-6_8.

Riguelle, F., Thomas, I., Verhetsel, A., 2007. Urban polycentrism: a measurable reality. The case of Brussels. Journal of Economic Geography 7, 193–215.

Robinson, W.S., 1950. Ecological correlations and the behavior of individuals. American Sociological Review 15, 351–357.

Rodrigue, J.P., Comtois, C., Slack, B., 2009. The geography of transport systems. Routledge, London.

Rosenbaum, A.M., Koenig, B.E., 1997. Evaluation of the Modelling Tools for Assessing Land Use Policies and Strategies. Technical Report. Technical report prepared for the transportation and market incentives group, United States Environmental Protection Agency.

Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric Regression. Cambridge University Press.

RVCoG, 1999. Travel Demand Modeling of Land Use Alternatives. Rogue Valley Council of Governments. http://www.rvcog.org/ftp/RVMPO/TPAU/.

Sarle, W.S., 1983. SAS Technical Report A-108, Cubic clustering criterion. Technical Report. SAS Institute Inc.

Schindler, M., Caruso, G., 2014. Urban compactness and the trade-off between air pollution emission and exposure: Lessons from a spatially explicit theoretical model. Computers, Environment and Urban Systems 45, 13 – 23. doi:http://dx.doi.org/10.1016/j.compenvurbsys.2014.01.004.

Schirmer, P., Zollig Renner, C., Mülcer, K., Axhausen, K., 2015. Land use and transport microsimulation in the canton of Zurich using UrbanSim, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 485–536.

Schuler, M., M, P., Dessemontet, P., 2009. Agglomération et aires urbaines dans l'espace transfrontalier Ain-Haute Savoie-Geneve-Vaud. Technical Report. Observation Statistique transfrontalier.

SDAG, 1995a. The Energy Yardstick: Using PLACES to Create more Sustainable Communities. San Diego Association of Governments. http://www.sandag.org/uploads/publicationid/publicationid_1082_3274.pdf.

SDAG, 1995b. The Euclid PLACE3S Revitalization Program. San Diego Association of Governments. http://www.sandiego.gov/planning/community/profiles/southeasternsd/pdf/euclidplaces3revitalizationprogram.pdf.

SDAG, 1996. The Energy Yardstick: Using PLACE3S to Create more Sustainable Communities. San Diego Association of Governments. http://www.smartcommunities.ncat.org/pdf/places.pdf.

Sener, I., Pendyala, R., Bhat, C., 2011. Accommodating spatial correlation across choice alternatives in discrete choice models: an application to modeling residential location choice behavior. Journal of Transport Geography 19, 294–303.

Sevcikova, H., Raftery, A., Waddell, P., 2007. Assessing uncertainty in urban simulations using bayesian melding. Transportation Research Part B 41, 652–659.

Seya, H., Yamagata, Y., Tsutsumi, M., 2013. Automatic selection of a spatial weight matrix in spatial econometrics: Application to a spatial hedonic approach. Regional Science and Urban Economics 43, 429 – 444. doi:http://dx.doi.org/10.1016/j.regsciurbeco.2013.02.002.

Shermann, L., 1994. Land use and transportation: envisioning regional sustainability. Transport Policy 5, 147–161.

Shmueli, G., 2011. To explain or to predict? Statistical Science 25, 289–310.

Shukla, V., Waddell, P., 1991. Firm location and land use in discrete urban space: a study of the spatial structure of Dallas-Fort Worth. Regional Sciences and Urban Economics 21, 225–253.

Simmonds, D., Feldman, O., 2005. Land-use/transport modelling with the DELTA package: progress and experience, in: Proceedings of the 9th International Conference on Computers in Urban Planning and Urban Management, University College London. http://www.davidsimmonds.com/files/resourcesmodule/@random4767836fd9d18/1200086520_cupum_paper_354.pdf.

Simmonds, D., Waddell, D., Wegener, M., 2013. Equilibrium versus dynamics in urban modelling. Environment and planning b: planning and design 40, 1051–1070.

SLRCG, 2003. Santee-Lynches Regional Infrastructure Plan. Santee-Lynches Regional Council of Governements. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1412/filename/1413.pdf.

Snickars, F., Johansson, B., Leonardi, G., 1983. Models as Integral Parts of Regional Information Systems: Experiences from Italy and Sweden. Technical Report. International Institute for Applied Systems Analysis. http://www.iiasa.ac.at/publication/more_WP-83-052.php.

Southworth, F., 1995. A technical review of urban land use-transportation models as a tool for evaluating vehicle travel reduction strategies. Technical Report. Oak Ridge National Laboratory. http://cta.ornl.gov/cta/Publications/Reports/ORNL-6881.pdf.

Sui, D., 1998. GIS-based urban modelling: practices, problems, and prospects. International journal of geographical information science 12, 651–671.

Tannier, C., Thomas, I., 2013. Defining and characterizing urban boundaries: a fractal analysis of theoretical and Belgian cities. Computer, Environment and Urban Systems 41, 234–248.

Tannier, C., Thomas, I., Vuidel, G., Frankhauser, P., 2011. A fractal approach to identifying urban boundaries. Geographical Analysis 43, 211–227.

TCRPC, 2002. Martin and St. Lucie Counties Regional Land Use Study. Treasure coast regional planning council. http://www.tcrpc.org/special_projects/regional_land_use_study/.

TCRPC, 2003. Choices for Our Future - Technical Memoranda to Tasks II-3.1 and II-3.2. Tri-County Regional Planning Commission. https://www.michigan.gov/documents/MDOT-PTD-RegionalTrends_114113_7.pdf.

Thill, J.C., 1992. Choice set formation for destination choice modeling. Progress in Human Geography 16, 361–382.

Thisse, J.F., Thomas, I., 2007. Bruxelles et Wallonie: une lecture en terme de geo économie urbaine. Reflets et Perspectives de la vie Economique 46, 75–93.

Thisse, J.F., Thomas, I., 2010. Bruxelles au sein de l'économie belge: un bilan. Reflets et Perspectives de la vie Economique 80, 1–18.

Thomas, I., Cotteels, C., Jones, J., Peeters, D., 2013. Revisiting the Extension of the Brussels urban agglomeration: new methods, new data... new results? e-Belgeo 1-2, 1–11. https://belgeo.revues.org/6074.

Thomas, I., Cotteels, C., Jones, J., Pholo Bala, A., Peeters, D., 2015. Spatial challenges in the estimations of LUTI models: some lessons from the sustaincity project, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 55–73.

Thomas, I., Zenou, Y., 1999. Ségrégation urbaine et Discrimination sur le marche du travail: le cas de Bruxelles, in: Catin, M., Lesueur, J., Zenou, Y. (Eds.), Emplois, Concurrence et Concentration Spatiale. Economica, Paris, pp. 105–127.

Thurstone, L., 1927. A law of comparative judgement. Psychological Review 34, 272–286.

Timmermans, H., 2003. The saga of integrated land use transport modeling: how many dreams before we wake up?, in: Moving through nets: The physical and social dimensions of travel, 10th International Conference on Travel Behaviour Research, Lucerne. pp. 1–35.

TJPDC, 2002. Building Livable Communities: Jefferson Area Eastern Planning Initiative. Thomas Jefferson Planning District Commission. `http://content.lib.utah.edu/cdm/ref/collection/FHWA/id/1420`.

Train, K., 2003. Discrete choice methods with simulation. Cambridge university press.

Turci, L., Pennec, S., Toulemon, L., Bringe, A., Baggio, R., Morand, E., 2015. Agent-based microsimulation of population dynamics, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 113–136.

TVFP, 2002. Treasure Valley Growth Scenario Analysis Final Report. Treasure Valley Futures Project. `http://www.compassidaho.org/documents/planning/studies/FinalTVFReport-GrowthScenarioAnaslysis.pdf`.

UN, 2014. World urbanisation prospects: the 2014 revision, highlights. United Nations, Department of Economic and Social Affairs, Population Division. `http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf`.

Van Hecke, E., Halleux, J.M., Decroly, J.M., Merenne-Schoumaker, B., 2009. Noyaux d'habitats et région urbaines dans une Belgique urbanisée. Technical Report. Monographie de l'enquète socio-économique 2001.

Van Malderen, L., Jourquin, B., Thomas, I., Vanoutrive, T., Verhetsel, A., Witlox, F., 2012. On the mobility policies of companies: what are the good practices? The Belgian Case. Transport Policy 21, 10–19.

Vandenbulcke, G., Steenberghen, T., Thomas, I., 2007. Accessibility indicators to place and transport. Technical Report. Politique scientifique fédérale BELSPO. http://www.belspo.be/belspo/organisation/publ/pub_ostc/AP/rAP02_en.pdf.

Vandermotten, C., Van Hamme, G., Kummert, P., Wayens, B., Williaert, D., Deboosere, P., Eggerickx, T., sanderson P-P, Kesteloot, C., de Maesschalck, D., 2010. La dynamique des quartiers bruxellois (1981-2006). Technical Report. Groupe d'étude de démographie appliquée. http://www.adt-ato.irisnet.be/sites/default/files/documents/Evolution_socio_economique_Rapport_final_FR.pdf.

Vanneste, D., Thomas, I., Goossens, L., 2007. Le logement en belgique. Technical Report. Monographie de l'enquète socio-économique 2001.

Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. Springer.

Verhetsel, A., Thomas, I., Beelen, M., 2010. Commuting in Belgian metropolitan areas. Journal of Transport and Land Use 2, 109–131.

Viegas, J.M., Miguel Martinez, L., Silva, E.A., 2009. Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. Environment and Planning B: Planning and Design 36, 625–643.

Vold, A., 2005. Optimal land use transport planning for the Greater Oslo area. Transportation Research Part A 39, 548–565.

Waddell, P., 2000. A behavioral simulation model for metropolitan policy analysis and planning: residential location and housing market components of UrbanSim. Environment and Planning B: Planning and Design 27, 247–263.

Waddell, P., 2002. UrbanSim: Modeling Urban Development for Land Use, Transportation and Environmental Planning. Journal of the American Planning Association 68, 297–314.

Waddell, P., 2011. Integrated Land Use and Transportation Planning and Modeling: Addressing Challenges in Research and Practice. Transport Reviews 31, 209–229.

Waddell, P., Borning, A., Noth, M., Freier, N., Becke, M., Ulfarsson, G., 2003. Microsimulation of Urban Development and Location Choices: Design and Implementation of UrbanSim. Networks and Spatial Economics 3, 43–67.

Waddell, P., Ulfarsson, G., Franklin, J., Lobb, J., 2007. Incorporating land use in metropolitan transportation planning. Transportation Research Part A 41, 382–410.

Wagner, P., Wegener, M., 2007. Urban land use, transport and environment models: experience with an integrated microscopic approach. disP 170, 45–56.

Wang, Y., Monzon, A., Di Ciommo, F., 2015. Assessing the accessibility impact of transport policy by land-use and transport interaction mode - the case of Madrid. Computers, Environment and Urban Systems 49, 126–135.

WAPC, 2000. 2025 Metropolitan Transportation Plan. Wilmington Area Planning Council. http://www.china-up.com:8080/international/case/case/197.pdf.

WAPC, 2003. Regional Transportation Plan 2025: Opening the Door to Change. Wilmington Area Planning Council. http://content.lib.utah.edu/utils/getfile/collection/FHWA/id/1415/filename/1416.pdf.

Wardman, M., 1988. A comparison of revealed preference and stated preference model of travel behaviour. Journal of Transport Economics and Policy 22, 71–91.

van Wee, B., 2015. Towards a new generation of land-use transport interactions models: a viewpoint. Journal of transport and land use 8, 3–10.

Wegener, M., 1994. Operational urban models: state of the art. Journal of the American Planning Association 60, 17–29.

Wegener, M., 1995. Current and future land use models, in: Shunk, G.A. (Ed.), Land Use Modeling Conference Proceedings, Texas Transportation Organization. http://www.spiekermann-wegener.com/pub/.

Wegener, M., 2004. Overview of Land-Use Transport Models, in: Hensher, A., Button, K. (Eds.), Transport Geography and Spatial Systems. Pergamon/Elsevier Science, Kidlongton, UK, pp. 127–146.

Wegener, M., 2011a. From Macro to Micro - How much micro is too much? Transport Reviews 31, 161–177.

Wegener, M., 2011b. The IRPUD model. Technical Report. Spiekermann & Wegener Stadt- und Regionalforschung. Dortmund. http://www.spiekermann-wegener.com/mod/pdf/AP_1101_IRPUD_Model.pdf.

Wegener, M., 2014. Land-use transport interactions models, in: Fischer, M.M., Nijkamp, P. (Eds.), Handbook of Regional Science. Springer-Verlag, Berlin, pp. 741–758.

Wegener, M., Furst, F., 1999a. Land-Use Transport Interaction: State of the Art. Technical Report. Institut fur Raumplanung. http://www.irpud.raumplanung.tu-dortmund.de/irpud/fileadmin/irpud/content/documents/publications/ber46.pdf.

Wegener, M., Furst, F., 1999b. Land use-transport interaction: state of the art. Technical Report. Institute of Spatial Planning, University of Dortmund. Dortmund. http://www.raumplanung.tu-dortmund.de/irpud/fileadmin/irpud/content/documents/publications/ber46.pdf.

Wegener, M., Gand, F., Vannahme, M., 1986. The time scale of urban change, in: Hutchinson, B., Batty, M. (Eds.), Advances in urban systems modelling. North-Holland, Amsterdam, pp. 175–197.

Wegener, M., Mackett, R., Simmonds, D., 1991. One city, three models: comparison of land-use/transport policy simulation models for Dortmund. Transport Reviews 11, 107–129.

Weidlich, W., Munz, M., 1990. Settlement formation. The Annals of Regional Science 24, 83–106. doi:10.1007/BF01579725.

White, R., Engelen, G., 1993. Cellular dynamics and GIS: modelling spatial complexity. Geographical systems 1, 237–253.

White, R., Engelen, G., Uljee, I., 2015. Modeling cities and regions as complex systems. The MIT Press, Cambridge, MA.

Wilhelmsson, M., 2002. Spatial Models in Real Estate Economics. Housing, Theory and Society 19, 92–101. doi:10.1080/140360902760385646.

Wilson, A.G., 1969. The use of entropy maximising models, in the theory of trip distribution, mode split and route split. Journal of Transport Economics and Policy , 108–126http://www.bath.ac.uk/e-journals/jtep/pdf/Volume_111_No_1_108-126.pdf.

Witlox, F., Jourquin, B., Thomas, I., Verhetsel, A., Van de Vijver, E., Van Malderen, L., Vanoutrive, T., 2011. Assessing and developing initiatives of companies to control and reduce commuter traffic. Technical Report. BELSPO. http://www.transport-research.info/sites/default/files/project/documents/20150324_121928_86085_ADICCT_FinRep_AD.pdf.

Wrigley, N., 1985. Categorical data analysis for geographers and environmental scientists. Longman, London and New York.

Yang, J.H., Ettema, D., Frenken, K., 2012. Modelling the emergence of spatial patterns of economic activity. ArXiv e-prints .

Ye, X., Konduri, K., Pendyala, R., Sana, B., Waddell, P., 2009. A methodology to match distributions of both household and person attributes in the generation of synthetic populations, in: 88th Annual meeting of the Transportation Research Board, Transportation Reasearch Board, Washington D.C.

Zenou, Y., 2009. Urban Labor Economics. Cambridge University Press.

Zhang, M., Kukadia, N., 2005. Metrics of urban form and the modifiable areal unit problem. Transportation Research Record 1902, 71–79.

Zhou, X., Lin, H., 2008. Spatial Weights Matrix, in: Shekhar, S., Xiong, H. (Eds.), Encyclopedia of GIS. Springer US, pp. 1113–1113. doi:10.1007/978-0-387-35973-1_1307.

Zollig Renner, C., Axhausen, K., 2015. A real estate development model with heterogenous agents, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 181–200.

Zollig Renner, C., Nicolai, T., Nagel, K., 2015. Agent-based land use and transport interaction modelling: state-of-the-art, in: Bierlaire, M., de Palma, A., Hurtubia, R., Waddell, P. (Eds.), Integrated transport and land use modelling for sustainable cities. Routledge/EPFL Press, Lausanne, pp. 9–32.

Zondag, B., Bok, M., Geurs, K., Molenwijk, E., 2015. Accessibility modeling and evaluation; the TIGRIS XL land-use and transport interaction model for the Netherlands. Computers, Environment and Urban Systems 49, 115–125.

Zondag, B., de Jong, G., 2011. The development of the TIGRIS XL model: a bottom-up approach to transport, land-use and the economy. Research in Transportation economics 31, 55–62.

List of Tables

List of Figures

Contents