Transport Reviews

ISSN: 0144-1647 (Print) 1464-5327 (Online) Journal homepage: http://www.tandfonline.com/loi/ttrv20

City delineation in European applications of LUTI models: review and tests

Isabelle Thomas, Jonathan Jones, Geoffrey Caruso & Philippe Gerber

To cite this article: Isabelle Thomas, Jonathan Jones, Geoffrey Caruso & Philippe Gerber (2017): City delineation in European applications of LUTI models: review and tests, Transport Reviews

To link to this article: http://dx.doi.org/10.1080/01441647.2017.1295112

Published online: 28 Feb 2017.

TRANSPORT REVIEWS, 2017
http://dx.doi.org/10.1080/01441647.2017.1295112

City delineation in European applications of LUTI models: review and tests

Isabelle Thomas[a], Jonathan Jones[a], Geoffrey Caruso [b] and Philippe Gerber[c]

[a]Center for Operations Research and Econometrics, Université catholique de Louvain, Louvain-la-Neuve, Belgique; [b]Institute of Geography and Spatial Planning, Université du Luxembourg, Maison des Sciences Humaines, Esch-sur-Alzette, Luxembourg; [c]Luxembourg Institute for Socio-Economic Research, Maison des Sciences Humaines, Esch-sur-Alzette, Luxembourg

ABSTRACT

This paper stresses the lack of attention paid to the geographical definitions of cities in LUTI models as one key detrimental aspect to transferring and generalising LUTI results. First, the argumentation develops from a meta-analysis of peer-reviewed publications about LUTI applications in European cities. We show that most authors do not assess findings against potential geographical biases. Second, theoretical simulations are conducted with *UrbanSim* applied to a synthetic urban area. By varying the geographical limits of the system and population endowments, our simulations confirm that the absence of control on city delineation weakens the results. Finally, the paper suggests methodological guidelines to improve the comparability of LUTI applications and push forward their theoretical agenda.

ARTICLE HISTORY
Received 18 July 2016
Accepted 9 February 2017

KEYWORDS
LUTI models; Europe; city delineation; urbansim; theoretical simulation

1. Introduction

Defining the physical and functional limits of a city is a difficult task and obviously depends on research and normative objectives. Each city has a different built-up footprint and a functional hinterland determined by many interacting actors and processes, themselves constrained by a particular physical geography, settlement history, economy, or land-use and transport planning culture. Defining cities morphologically and functionally is a key debate in urban geography literature continuously revived by evolving location and transport behaviour and by new data (big data, crowd-sourced transport information, etc.). It is less so in *Land-Use and Transport Interaction* (LUTI) literature. Rather, since the beginning of LUTI models, researchers and consultants have taken a very pragmatic view on the geographical delineation of the urban systems.

In the early stage of LUTI models, Lowry (1964) reflected upon finding relevant spatial units for his model and opted for a gridded representation of data. In terms of the limits of the study area; however, his choice was entirely data driven: "Since original field work was out of the question, the model had to be accommodated to existing data-files" (Lowry, 1964, p. 55). He then addressed the size of his case study in terms of total surface,

CONTACT Isabelle Thomas isabelle.thomas@uclouvain.be Center for Operations Research and Econometrics, Université catholique de Louvain, Voie du Roman Pays 34, 1348 Louvain-la-Neuve, Belgique

population and jobs and provided a map of the delineation of the area. He also stressed the importance of having half of the total surface available for future growth. The link between the selected study area and the goal of the model was clear, while today's literature seems rather silent on these points.

While practical concerns are perfectly understandable, there is a risk of undermining the effects of choosing the limits of an urban system. In most disciplines, assessing the effect of system boundaries is important to provide certainty and robustness to model findings. This has long been an issue for geographers, either from theoretical or empirical perspectives (see, e.g. Berry & Lamb, 1974; Griffith, 1983; Hall, 2007; Kwan, 2012; Openshaw, 1984 for the former, and Chakraborty, Wilson, & bin Kashem, 2015; Jones, Peeters, & Thomas, 2015a; Raciti, Hutyra, Rao, & Finzi, 2012 for the latter). In urban and transport research, it is of foremost importance as soon as one intends to contribute generalisable scientific knowledge about the functioning of city regions. It is especially true for LUTI models since they involve complex processes that are more and more geographically detailed and disaggregate in terms of agents, and since they include many nonlinear interactions and feedbacks across scales.

Ideally, contributed knowledge should be easily detachable from a specific dataset for one particular region, and should also be easily transferable across case studies. Specific local actors and processes, especially the decisions of urban planners and policy-makers, obviously make such generalisation difficult. Nevertheless, identifying this local context precisely requires transparency in how the geographical limits are selected and how this choice impacts the spatial distribution and variability of data and whether or not it affects results. In this paper we show how the size and extent of cities are dealt with in recently published LUTI literature. We also show how they affect the results of an LUTI model applied to a synthetic metropolitan area.

Land-Use and Transportation Interaction (LUTI) models are derived from the classical four-step model. The first-generation of models, coming shortly after Lowry (1964), led to numerous applications in the USA (e.g. MEPLAN, see Echenique et al., 1990; Hunt & Simmonds, 1993; or TRANUS, see de la Barra, 1989) due to federal regulations requiring land-use impacts of new transport infrastructure to be assessed. Their spatial diffusion to Europe is more recent and scarce. Today, LUTI models still appear to be mostly a pragmatic integration of bits of land-use within transport models rather than the opposite, and therefore are weakly connected to urban land-use theory (Alonso, 1964 and subsequent literature), especially regarding urban dynamics (Anas, 2013b).

As geographers and cartographers know for long, a city is not simply a dot on a map but is characterised by extent, morphology, and mass (e.g. population) (Batty, 2008; Slocum, McMaster, Kessler, & Howard, 2005). Moreover, each city has a different footprint determined by complex interacting actors and processes, including its geography, history, and governance structure (e.g. Abdel-Rahman & Anas, 2004; Batty, 2005; Derycke, Huriot, & Pumain, 1996; Parr, 2007; Tannier & Thomas, 2013). Furthermore, there is no international agreement about the delineation of functional urban regions (NUREC, 1994) and various cut-off values are used for example to delineate commuting basins (e.g. Cheshire & Gornostaeve, 2002; Dujardin, Thomas, & Tulkens, 2007; Thomas, Cotteels, Jones, & Peeters, 2013). Cities are interconnected, hierarchically organised (Pumain, 2006) and hence difficult to isolate. It is especially difficult to delineate a city because built-up surfaces dilute farther and farther away from the city centre within rural areas, thus blurring the morphological and functional limits of the city (Caruso, Peeters, Cavailhès, & Rounsevell, 2007, 2011). This problem becomes even more complex when the polycentric nature of the urban system is considered. These urban processes are a source of complex and multi-faceted mobility behaviours (e.g. Cervero, 2002; Handy, Boarnet, Ewing, & Killingsworth, 2002).

When selecting a city for an LUTI application, the larger the study area, the larger the risk of including functional or morphological parts of other cities. A different delineation can therefore change the nature of the studied urban system and automatically determine important parameters of the model, including transport outcomes, as well as the level of system response to scenarios. This already was a point in Lee's (1973) requiem on large-scale models, emphasising wrongheadedness when implementing gravity processes at all scales. Bypassing the discussion on the size and extent of cities, LUTI modellers may have slowed down the integration in urban theory of their findings related to particular cases and geographies. After almost 50 years of applications, the theoretical and legitimate promises of a fully interacting land-use and transport system have then not been fully met. While mature in practical terms, the field still seems to lack the capacity to contribute generalisable scientific knowledge about the functioning of city regions, especially because of difficulties in case studies comparability and full transparency in application. On the other extreme, the standard urban economic literature, after Alonso, while informing clearly on generic policies and optimal instruments (cordon, taxes, etc. to tackle sprawl, congestion or negative externalities; e.g. Brueckner, 2001; De Borger, Proost, & Van Dender, 2008) stays far from empirical validation beyond stylised facts, and real case calibration and implementation are rare. Hence urban geography and policy, or planners cannot easily capitalise on transport and land-use models. Our statement here is somehow a rejoinder to Saujot, de Lapparent, Arnaud, and Prados (2016) who emphasise a gap between theory and end-users. Our viewpoint, however, is rather than the outcome of LUTI implementations may not be sufficiently general and robust to transfer between cases with a different geography.

Paulley and Webster (1991) summarised a comparative study of six applied LUTI models. Among other things, they concluded that the size and characteristics of the city to which the models are applied influence the results. Acheampong and Silva (2015) have recently conducted a comprehensive review of the entire LUTI field and challenges but they do not address the effect of the city definition on results. Our paper questions whether Paulley and Webster's lesson on city definition has been taken up in more recent research. We first perform a meta-analysis of the articles published over the last 25 years where LUTI model applications are described (Section 2). Second, we run simulations on a theoretical study area using a well-established LUTI model (*UrbanSim*) in order to test generic effects of urban delineation on LUTI outcomes (Section 3). Finally, we propose recommendations to mitigate the problems related to area delineation (section 4).

2. LUTI models applied to European cities: a meta-analysis

2.1. Corpus

Our search for scientific publications was restricted to *peer-reviewed journals* with impact factor and published between 1990 and June 2016. All other types of documents, such as proceedings, working papers, chapters in books, and reports were not considered here.

This choice guarantees quality, although we are conscious that it may slightly bias final interpretation. We particularly looked for applications of LUTI models in using *Google Scholar*, *Scopus* and *ResearchGate*. Given that "LUTI models" are close to other transportation and land-use models, our decision to include a paper in our corpus was based on whether the author(s) themselves stated their model was an LUTI or not.

The analysis is limited to European cities because their urban structure differs from Asia or the USA (see e.g. Brueckner, Thisse, & Zenou, 1999; Bertaud & Malpezzi, 2003; Hohenberg & Lees, 1986). Moreover, conversely to the USA where the same methodology is used to delineate all urbanised areas (see Federal Register, 2000, 2002), criteria and thresholds vary from one country to another in Europe, making our problem more acute (Cörvers, Hensen, & Bongaerts, 2009; Dujardin et al., 2007). Table 1 reports the information collected for each case study.

2.2. A small number of applied academic papers

Despite the existence of a large number of LUTI platforms, academic papers obviously do not favour empirical analysis (confirming Wegener, 2011) but rather methodological developments. Only *21 applied scientific papers* were found. Interestingly it seems that

Table 1. Cities reported in the meta-analysis.

City	Authors	Journal[a]	Year	Size[b]	Population[c] (10^6)	Area[c] (km^2)	Map[b]
Bilbao	Echenique et al	TR	1990	N	(0.95)		N
Dortmund	Echenique et al	TR	1990	N	(0.58)		N
Dortmund	Mackett	TR	1990	Y	1.1	833	N
Leeds	Echenique et al	TR	1990	N	(0.72)		N
Leeds	Mackett	TR	1990	Y	0.5	164	N
Dortmund	Wegener et al	TR	1991	Y	2.3		Y
Stockholm	Anderstig et al	Transp.	1992	N	(2.1)		N
Bilbao	Burgos	EPb	1994	Y	1		N
Naples	Hunt	EPb	1994	N	(4.4)		Y
The Netherlands	Eradus et al	JTG	2002	Y	16		Y
Edinburgh	May et al	TRR	2005	Y	2.7	2305	N
Leeds	May et al	TRR	2005	Y	2.1	559	Y
Oslo	Vold	TR-A	2005	Y	0.95		N
Dortmund	Wagner et al	disP	2007	Y	2.6		Y
Brussels	Patterson et al	JUPD	2010	Y	2.9	4361	Y
Brussels	Patterson et al	EPa	2010	Y	2.9	4361	Y
Lausanne	Patterson et al	EPa	2010	Y	0.277	200	Y
Lyon	Patterson et al	JUPD	2010	Y	1.6	3325	Y
Rome	Di Zio et al	JUPD	2010	Y	2.6	1500	Y
Paris	Anas	EPb	2013a	N	(12)	(762)	Y
London	Batty et al	EPb	2013	Y	14		Y
Santander	Coppola et al	JUPD	2013	Y	0.28		Y
Madrid	Wang et al	CEUS	2015	Y	6.5	8030	Y
Thessaloniki	Pozoukidou	Spatium	2014	N	(1.08)	(1455)	Y
Besançon	Bonin & Tomasoni	IJTr	2015	N	0.179	(222)	Y
Madrid	Guzman et al	CSTP	2015	Y	6.5	8000	Y
The Netherlands	Zondag et al	CEUS	2015	N	16		N

[a]TR = Transportation review; Transp. = Transportation; EPa/b = Environment and Planning A/B; JTG, *Journal of Transport Geography*; TRR, *Transportation Research Record*; TR-A, *Transportation Research – A*; disP = disP – The planning review; JUPD = *Journal of Urban Planning and Design*; CEUS, *Computer, Environment, and Urban Systems*; IJTr, *International Journal of Transportation*; CSTP, *Case Studies on Transport Policy*.
[b]Y = included in the paper, N otherwise.
[c]Value between bracket not mentioned; estimations by other means.

there are virtually no applied peer-reviewed publications before our cut-off date of 1990, corresponding more or less to the ISGLUTI report by Webster and Dasgupta (1991) and a special issue of Transport Reviews in 1990. This may be seen as the impact of Lee's requiem (Jones, 2016) but may also demonstrate an early focus on model development rather than on capitalising knowledge from applied cases.

The small total number of published applications can be seen as a result of several causes. First, it can partly be due to the fact that recent LUTI models require micro-data that are often difficult to collect (availability, privacy reasons) and costly to analyse. This problem was already mentioned by the users of the very first LUTI platforms (e.g. Burgos, 1994; Hunt, 1994) and is still up to date despite the emergence of "big-data" and concomitant increase in computational capacities. Second, the variety in disciplinary backgrounds of the authors adds difficulties in the speed of execution of the work. Synthesis of a multidisciplinary approach in a compact scientific paper within the 3–4 years of a project's timeline is difficult, and automatically leads to some drastic choices. Publishing both methodological innovations and empirical results in a single paper is indeed highly challenging, due to the format of most journal articles. Third, the small number of published applications also pertains to the type of funding: a research project financed by/to a consulting firm will clearly lead to operational contributions and/or technical reports rather than academic papers. Fourth, most papers are published in transportation journals, while urban studies and geography journals have welcomed such work more recently. We also suspect a high general reluctance of journals to accept papers with an application focus, Yet, this seems to be less the case in the most recent years of our corpus, which is perhaps in line with the release of new dedicated journals such as *Journal of Transport and Land Use* or *Case Studies in Transport Policy*.

2.3. A limited set of urban geographies with unclear limits

Within the 21 surveyed papers, 5 analyse more than one city, leading to a total of 27 case studies (Table 1 and Figure 1). Dortmund, Leeds, Brussels, Bilbao, and Randstad Holland are studied more than once. This repetition in case studies can certainly be related to time-consuming data collection and parameter hungry implementation steps. Examples and prepared dataset are *used and re-used* for scale economies, which are also important for academic authors. We can also relate the lack of diversity in case studies to difficulties in raising awareness within the community of planning practitioners and policy-makers of the potential benefits to having an LUTI model developed for their city and ready to simulate scenarios.

Two papers deal with a *network of cities*. They correspond to local transport planning problems embedded within a broader application of TIGRIS in The Netherlands and Randstad Holland (Eradus, Shoemakers, & van der Hoorn, 2002; Zondag, Bok, Geurs, & Molenwijk, 2015). All other case studies (25 out of 27) refer to *isolated urban area*(s). Their delineations are not well documented: most papers do not even include a paragraph or a map devoted to the description of the study area. Official metropolitan areas are often used, but authors neither mention the exact limits, nor the year considered, thus accepting with no critical viewpoint that metropolitan boundaries are adequate. The delineation is governed by policy/administrative reasons, by the agencies supplying the data and/or

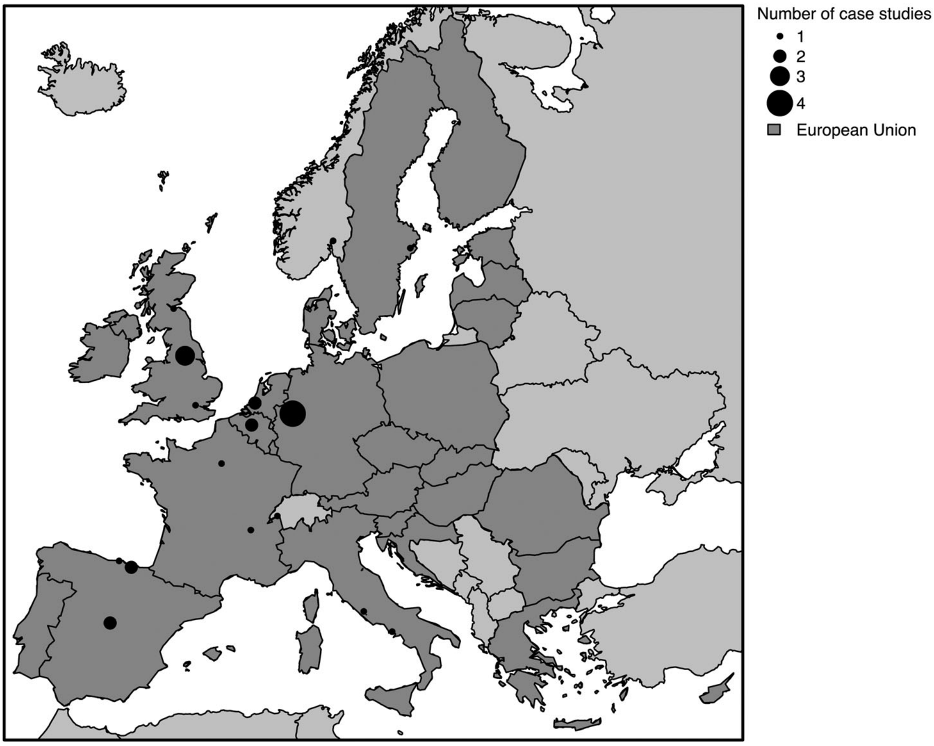

Figure 1. LUTI models' applications in Europe: location and frequency.

by the researchers assembling the available data. Mackett (1990) raises that problem without solving it.

Some studied areas are much larger than others, and include rural landscapes (e.g. Paris Ile-de-France, officially defined by the National Statistical Office). In other cases, they encompass parts of the functional or morphological agglomeration of neighbouring cities. This may radically affect accessibility measures, the type of transport practices and actors considered, and call for different types of scenarios. For example, the delineation of Brussels used in the *SustainCity* project includes parts of the cities of Antwerp and Ghent and the entire hinterland of Leuven, while the scenarios relate to intra-urban structures only (Jones, 2016; Thomas et al., 2013). This also questions intercity relationships and the closure of the model, i.e. what is "the Rest of the World", its relative attractiveness, and how does it interact with the transport components of the model? In none of our 27 case studies is the use of external zones explicitly stated, while it is of particular importance to discuss how the Rest of the World relates to the modelling framework.

Surprisingly enough, in 25% of the reviewed papers, the studied urban area does not even have a dimension (no population, no surface). For the other 75%, the size of the studied areas varies between 0.2 (Besançon) and over 12 Mio inhabitants (Paris, London). We know that transport infrastructures, modal choices, friction of distance and land-use realities depend upon that size. Can we really expect parameters taken from the literature (as is often the case) to be relevant for other applications? Especially when the geographical context in which they have been estimated appear to be very different and, actually, are not even described? How can we then generalise the land-use or transport processes, the parameters estimates, or the implications of a policy scenario while contexts are very diverse and the number of cases limited?

3. Simulations on a synthetic metropolitan area

Simulations are performed on a theoretical study area using *UrbanSim* (see Waddell, 2011) in order to test generic effects of urban delineation on LUTI outcomes. The synthetic geography is chosen in order to mimic a monocentric city on a flat plain with a circular functional hinterland (commuting basin) near a second city. By varying the geographical limits of the system, subparts of this hinterland are either cut out of the model or intermingle with the second centre. Population endowments are also varied so as to generalise for a system of two equal cities or a system of one city with a peripheral subcentre.

3.1. Synthetic geography and model set-up

We consider a study area made of a rectangular grid of 60 horizontal and 25 vertical cells. Each cell represents a basic spatial unit, i.e. a *zone* of 1×1 km.[1] The study area is divided into three regions: the western and eastern regions are identical in size (625 zones, or 25×25 km) and possess a CBD (called West and East) in their centre and the remaining central region (250 zones, or 10×25 km) is a suburban area, as shown in Figure 2. We simulate the evolution of this metropolitan area for seven different cases: the *Complete* study area and six gradually smaller geographical *Extents,* named after the number of horizontal cells from the Western border. They result in a progressive inclusion of the East CBD into the studied area. Each *Extent* constitutes a closed world, meaning that the model does not represent commuting or migration flows coming from the "Rest of the World". This is a methodological limit related to the coupling of *UrbanSim* and *MATsim*, which does not permit such interaction. Our choice is nevertheless, consistent with our review objectives and with

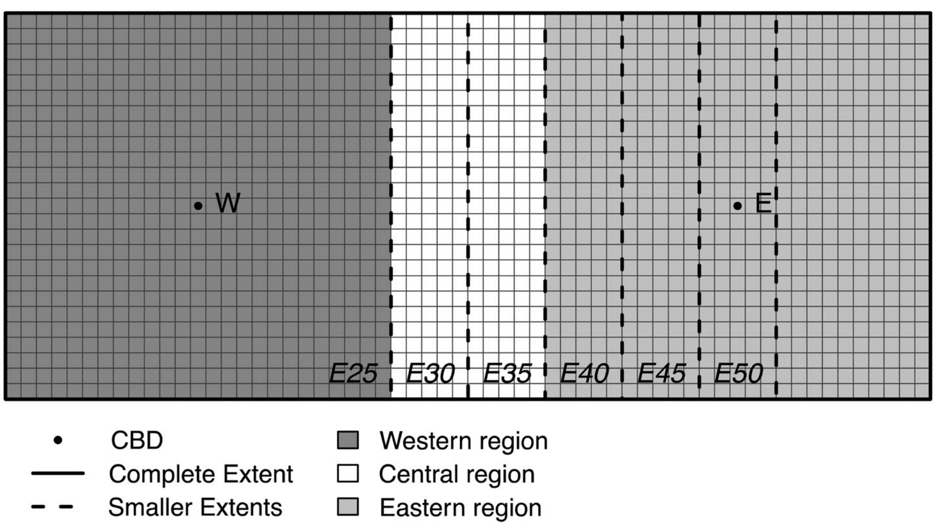

Figure 2. Structure of the theoretical study area and variation of the *Extents*.

recent real case LUTI applications (e.g. Cabrita et al., 2015; de Palma, Picard, & Motamedi, 2015; Schirmer, Zollig Renner, Mülcer, & Axhausen, 2015). Furthermore, the *Complete* extent constitutes here an isolated world, therefore. For any smaller *Extents*, the Rest of the World is thus limited to the non-included portion of the *Complete* study area.

In addition to changing the area extents, three kinds of relative population endowment are simulated: *equal-sized* CBDs, a *small West* CBD (West CBD half the size of the East CBD for households and jobs) and a *large West* CBD (West CBD twice the size of the East CBD). We end up with 7×3, i.e. 21 cases and will use the *Complete* cases as reference.

UrbanSim (see Noth, Borning, & Waddell, 2003; Waddell, 2000; or Hunt, Miller, & Kriger, 2005 for methodological details, and Section 3.2) is used to simulate the location of agents and is interfaced with an external travel model, *MATSim* (Nagel et al., 2008; Nicolai & Nagel, 2015). For each zone, the model provides the number of households, the number of jobs, commuting travel time and real estate prices. *UrbanSim* represents agents (households and jobs) in a disaggregated way. All agents are *homogenous*, meaning that all households have identical characteristics and that all jobs belong to the same employment sector.

The simulation period is 10 years, and a linear growth of 1% for both households and jobs is assumed per annum. Each iteration of *UrbanSim* accounts for one year while, for computational reasons, *MATSim* is run with an interval of three years. Each combination of CBDs' *size* and area *Extent* is simulated 30 times to cope with the stochastic nature of the model. All results presented are averaged values across these 30 simulations.

Since the landscape is featureless, the initial number of jobs and households per zone is a function of the Euclidean distance to both CBDs (see Appendix 1). The main inputs of the model are presented in Figure 3.

The initial conditions and characteristics of the zones are constant across all delineations of the study area *(Extents)*, i.e. each zone starts with the same number of agents and real estate prices in t_0. There are two exceptions: first, the Euclidean distance to the CBD is computed to the closest CBD for the *Complete* case, and to the Western CBD for all other cases; second, the travel model (see next section) estimates car accessibility to jobs independently for each *Extent*.

3.2. Econometric estimations

The calibration of *UrbanSim* requires the estimation of different econometric components, hereafter referred to as "sub-models".

Both household and job location models are a key components of the model system to forecast the future distribution of jobs and households. They rely on a multinomial logit model with random sampling of alternatives (see Waddell et al., 2003 for further details). Since the synthetic city is assumed to be in a steady state, both sub-models are estimated using a stratified sample of 10% of the agents per zone. The specification of these sub-models has been kept simple in order to minimise the number of feedback loops. Three variables are included for the households: car accessibility to jobs, residential prices and Euclidean distance to the CBD. Including the latter leads to the expected sign for the estimates of car accessibility and real estate prices (i.e. positive and negative) while keeping the model as simple as possible (the distance to CBD is constant over time). Non-home-based jobs only depend on car accessibility to jobs and on real estate prices of non-residential buildings. These are standard variables considered in economic geography.

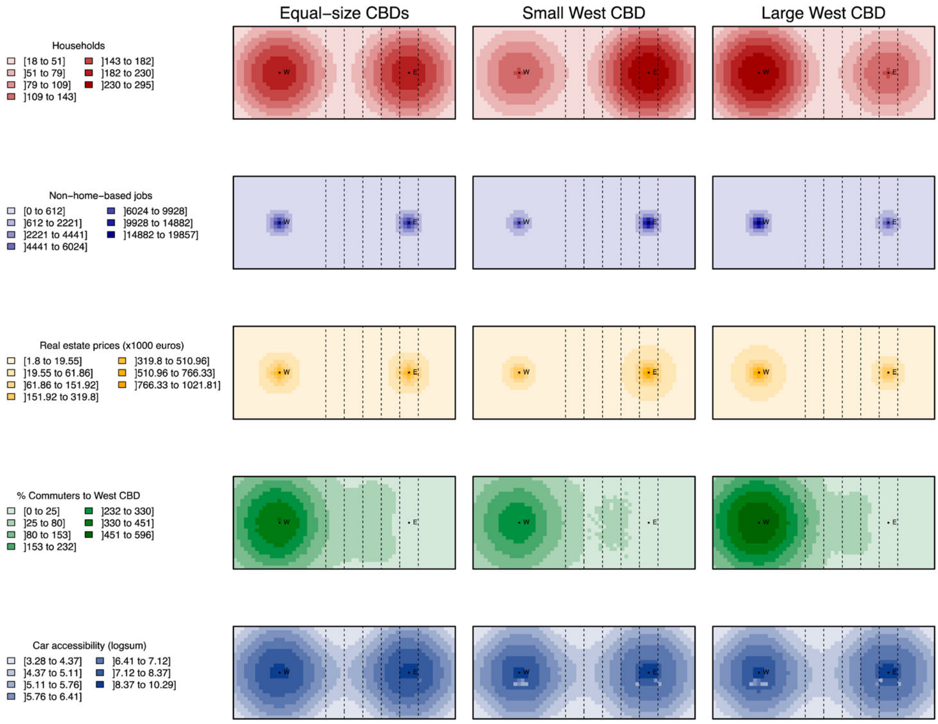

Figure 3. Model inputs in t_0.

MATSim computes the evolution of the car accessibility over time (Nicolai & Nagel, 2015), while the real estate price sub-model, using a log-linear form, updates real estate prices at the end of each *UrbanSim* iteration.

Despite the low predictive power of the household location model (Table 2), explanatory variables for both location models have expected signs. As expected as well, high utilities are found close to the CBD for all *Extents*. Parameter estimates for the real estate price sub-model (Table 3) are estimated using the initial situation of the synthetic city (see Appendix 1); they show a positive effect of both population and job densities.

Across the different *Extents*, we find important variations of the parameter estimates, confirming the already reported Modifiable Areal Unit Problems effects (see, e.g. Arauzo-Carod & Manjón-Antolín, 2004; Fotheringham & Wong, 1991; Jones et al., 2015a; Jones, Thomas, & Peeters, 2015b). A second issue is endogeneity: real estate prices may depend upon unobserved characteristics, not included in the real estate price sub-model but correlated with independent factors used in the location choice sub-models. The synthetic case study, however, avoids this potential bias since real estate prices only depend on population and jobs' densities, as indicated in Appendix 1 and showed by the adjusted R^2 nearly above 0.99 in Table 3.

In our experiments, these variations mean that the arrival of one new household in one particular zone for example will induce a different growth of real estate price in that zone. Consequently, an identical change of population or job density will lead to different variations of the utility level of this particular zone from one *Extent* to another. Given the

Table 2. Parameter estimates of the agents' location choice sub-models.

Extent	Variable	Households			Non-home-based jobs		
		Equal-size CBDs	Small West CBD	Large West CBD	Equal-size CBDs	Small West CBD	Large West CBD
Complete	Prices	−4.26e-06*	−3.90e-07	−1.63e-06***	−1.17e-06***	−3.87e-09	2.78e-06***
		(2.07e-06)	(3.76e-07)	(3.76e-07)	(7.31e-08)	(5.14e-08)	(3.85e-08)
	Car accessibility	0.053 (0.03)	0.07*** (0.02)	0.008 (0.021)	1.82*** (0.01)	1.38*** (0.009)	0.96*** (0.007)
	Dist. to CBD	−0.005 (0.01)	0.02** (0.006)	−0.007 (0.006)			
	AIC	10,6577	106,473	106,471	11,0236	11,3000	117,307
	LR	0.0002	0.0001	0.0002	0.46	0.45	0.43
	Observations	15,670	15,654	15,654	30,418	30,374	30,374
E50	Prices	−2.28e-06***	−7.10e-07	−6.15e-07	5.06e-06***	3.07e-06***	6.94e-06***
		(5.47e-07)	(5.81e-07)	(3.66e-07)	(3.88e-08)	(3.47e-08)	(2.85e-08)
	Car accessibility	0.022* (0.011)	0.045*** (0.013)	0.006 (0.005)	0.58*** (0.004)	0.86*** (0.005)	0.08*** (0.002)
	Dist. to CBD	−3.21e-06 (0.0007)	−0.001 (0.001)	−0.002 (0.001)			
	AIC	88,683	82,967	94,288	12,5750	12,0787	132,646
	LR	0.0002	0.0002	0.0002	0.37	0.38	0.34
	Observations	13,039	12,199	13,863	29,340	28,970	29,641
E45	Prices	−1.91e-06**	−1.57e-06	−1.14e-08	8.05e-06***	3.03e-06***	7.74e-06***
		(5.93e-07)	(1.30e-06)	(4.10e-07)	(5.19e-08)	(1.33e-07)	(3.35e-08)
	Car accessibility	0.006 (0.012)	0.01 (0.03)	0.003 (0.005)	0.22*** (0.005)	1.26*** (0.01)	−0.04*** (0.002)
	Dist. to CBD	0.002 (0.001)	0.004 (0.002)	0.0007 (0.001)			
	AIC	75,397	65,394	85,413	69,470	47,711	86,513
	LR	0.0002	0.0002	0.0002	0.37	0.39	0.39
	Observations	11086	9615	12558	16287	11589	20922
E40	Prices	−2.67e-06***	−3.06e-06**	−1.14e-06	6.55e-06***	2.81e-06***	5.91e-07***
		(6.62e-07)	(1.08e-06)	(1.10e-06)	(5.55e-08)	(1.36e-07)	(5.48e-08)
	Car accessibility	0.012 (0.014)	0.008 (0.03)	0.015 (0.024)	0.38*** (0.006)	1.14*** (0.01)	1.42*** (0.01)
	Dist. to CBD	−0.0007 (0.003)	0.0006 (0.005)	−0.002 (0.003)			
	AIC	64,195	50,620	77,736	63,575	40,042	74,505
	LR	0.0002	0.0004	0.0003	0.38	0.42	0.45
	Observations	9439	7444	11,430	15,244	10,226	20,218
E35	Prices	−1.0e-06	−1.006e-05***	−3.35e-06**	−1.09e-06***	2.46-e-06***	7.04e-06***
		(1.4e-06)	(1.6e-06)	(1.05e-06)	(1.0e-07)	(1.36e-07)	(3.28e-08)
	Car accessibility	0.003 (0.045)	0.13*** (0.03)	0.008 (0.006)	1.80*** (0.02)	1.21*** (0.01)	−0.029*** (0.002)
	Dist. to CBD	−0.006 (0.01)	0.0009 (0.007)	−0.023** (0.008)			
	AIC	57,231	41,679	72,730	54,789	40,449	86,061
	LR	0.0002	0.0004	0.0003	0.47	0.41	0.37
	Observations	8415	6129	10,694	15,209	10,185	20,189
E30	Prices	−2.89e-06	−7.09e-07	−6.52e-07	−1.42e-06***	2.93e-06***	6.92e-06***
		(2.83e-06)	(3.86e-06)	(2.12e-06)	(1.03e-07)	(1.29e-07)	(3.29e-08)

	Car accessibility	0.05 (0.04)	0.013 (0.034)	0.001 (0.006)	1.88*** (0.02)	1.16*** (0.01)	−0.008** (0.003)
	Dist. to CBD	0.001 (0.02)	0.014 (0.017)	−0.015 (0.019)			
	AIC	53,287	37,136	69,320	54,389	40,104	86,060
	LR	0.0002	0.0006	0.0002	0.47	0.42	0.37
	Observations	7835	5462	10,192	15,209	10,185	20,189
E25	Prices	−1.20e-06 (4.63e-06)	−7.81e-06 (7.89e-06)	−6.3e-06* (3.1e-06)	−1.92e-07* (8.79e-08)	3.0e-06*** (1.3e-07)	6.92e-06*** (3.27e-08)
	Car accessibility	0.11* (0.044)	0.006 (0.03)	0.0004 (0.007)	1.59*** (0.01)	1.14*** (0.01)	−0.012*** (0.003)
	Dist. to CBD	0.034 (0.035)	−0.026 (0.038)	−0.06* (0.03)			
	AIC	49,338	33,725	64,780	56,582	40,141	86,126
	LR	0.0003	0.0006	0.0002	0.45	0.42	0.37
	Observations	7255	4960	9525	15,209	10,185	20,189

Note: Between brackets: standard deviation; $\alpha \leq 0.05$*, 0.01**, and 0.001***; LR, Likelihood Ratio; AIC, Akaike Information Criterion.

Table 3. Parameter estimates of the real estate price sub-model.

		Non-residential buildings		
Extent	Variables	Equal-size CBDs	Small west CBD	Large west CBD
Complete	Constant	6.21*** (0.08)	5.61*** (0.06)	5.60*** (0.06)
	(Log) Job density	0.35*** (0.004)	0.32*** (0.004)	0.32*** (0.004)
	(Log) Pop. Density	0.39*** (0.01)	0.51*** (0.01)	0.51*** (0.01)
	R^2	0.95	0.95	0.95
	Observations	1500	1500	1500
E50	Constant	6.27*** (0.08)	5.71*** (0.07)	5.72*** (0.07)
	(Log) Job density	0.36*** (0.005)	0.34*** (0.005)	0.33*** (0.004)
	(Log) Pop. Density	0.38*** (0.0176669)	0.48*** (0.01)	0.49*** (0.01)
	R^2	0.95	0.95	0.95
	Observations	1250	1250	1250
E45	Constant	5.87*** (0.08)	5.52*** (0.07)	5.41*** (0.07)
	(Log) Job density	0.32*** (0.005)	0.30*** (0.004)	0.31*** (0.005)
	(Log) Pop. Density	0.46*** (0.01)	0.53*** (0.01)	0.55*** (0.01)
	R^2	0.95	0.95	0.96
	Observations	1125	1125	1125
E40	Constant	5.86*** (0.08)	5.54*** (0.08)	5.62*** (0.07)
	(Log) Job density	0.33*** (0.005)	0.31*** (0.005)	0.33*** (0.005)
	(Log) Pop. Density	0.46*** (0.01)	0.52*** (0.01)	0.51*** (0.01)
	R^2	0.96	0.94	0.96
	Observations	1 000	1 000	1 000
E35	Constant	5.98*** (0.09)	6.14*** (0.1)	5.68*** (0.07)
	(Log) Job density	0.34*** (0.006)	0.34*** (0.006)	0.33*** (0.005)
	(Log) Pop. Density	0.44*** (0.01)	0.39*** (0.02)	0.50*** (0.01)
	R^2	0.96	0.95	0.96
	Observations	875	875	875
E30	Constant	6.21*** (0.11)	6.38*** (0.12)	6.10*** (0.11)
	(Log) Job density	0.35*** (0.006)	0.35*** (0.006)	0.34*** (0.006)
	(Log) Pop. Density	0.39*** (0.02)	0.34*** (0.02)	0.42*** (0.02)
	R^2	0.95	0.95	0.96
	Observations	750	750	750
E25	Constant	6.94*** (0.16)	6.93*** (0.16)	7.13*** (0.16)
	(Log) Job density	0.37*** (0.007)	0.37*** (0.007)	0.38*** (0.007)
	(Log) Pop. Density	0.26*** (0.03)	0.23*** (0.03)	0.24*** (0.03)
	R^2	0.94	0.94	0.94
	Observations	625	625	625

Note: Between brackets: standard deviation; all parameters significant at $\alpha \leq 0.001$.

sequential nature of *UrbanSim*, there will be a snowball effect on the location choice models, and an uncertain final effect on all other model outputs.

In addition to the household and job location models, econometric estimations are made, respectively for the location of home-based jobs as a function of population density, and for the location of future real estate projects as a function of real estate prices and car accessibility (see Jones, 2016 for further details).

3.3. Predictions

As a calibration step, we first examine the evolution of land uses, prices and commuting times between t_0 and t_{10} (Table 4). The order of magnitude of the values per zone in t_{10} reflects the exogenous growth of the population. The largest differences with t_0 are observed for jobs and can be explained by the small number of jobs in many peripheral zones (see input in Figure 3). Moreover, most of the changes take place near the CBDs (see negative Pearson correlations in Table 4), which is consistent with our previous econometric estimates (see Tables 2 and 3).

Table 4. Calibration for equal-size CBDs (*Complete Extent* in t_{10} and evolution between t_0 and t_{10}).

Indicator	Value in t_{10} Min	Mean	Max	r^a
Households	21,93	110,40	243,97	−0.97
Jobs	0,00	211,50	16,371,00	−0.38
Prices	2 109	21 382	820 021	−0.47
Commuting times	0,00	60,96	119,19	0.86
	Evolution $t_0 - t_{10}$ (in %) Min	Mean	Max	
Households	9,44	10,56	14,13	−0,97
Jobs	−100,00	31,45	100,00	−0,15
Prices	3,59	6,75	10,48	−0,47
Commuting times	NA	NA	NA	NA[b]

[a]r = Pearson correlation between the Euclidean distance to the closest CBD and indicator (column 1); all correlations significant for $\alpha < .001$.
[b]Commuting times are estimated by *MATSim* at the end of t_0 and therefore not available for this time-step.

Four indicators are used to compare the land-use and transport outcomes of the LUTI model across the different *Extent*s: the final number of (1) households and (2) non-home-based jobs per zone[2] 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* case is used as reference, and relative differences (in %) are computed for each *Extent*: a negative difference means larger final values than for the *Complete* case, and vice-versa. Deviations from the *Complete* area are given both for the regions (Western, Central, and Eastern, in Table 5) and for the zones (see maps in Figures 4–7). The number of agents varies from one *Extent* to another. Therefore, we focus here on the spatial structure of the differences rather than discussing their absolute level.

Households – Variations per region and zones between *Complete* area and small *Extents* are limited (Table 5 and Figure 4). Since their magnitude is lower than the inter-runs variations (which vary from 0.58 to 1.27%) they can be considered as noise with no specific spatial pattern. Changing the size of the CBDs slightly influences these variations (increasing the size of the Western CBD leads to a decrease in variation). Note that the growth of the number of households between t_0 and t_{10} is larger near the CBDs for all *Extents*.

Jobs – Large differences are observed (Table 5 and Figure 5), and they are significantly larger than inter-run variations. A strong spatial structure emerges, with a higher concentration of jobs near the CBD for all smaller *Extents* (negative differences with the *Complete Extent*), and a lower one in the "dense suburbs". From *Extent E35* to *E25*, this denser suburb shrinks to a ring with limited width (see Figure 5 for details). Due to the concentration of jobs close to the CBDs, no variation is observed for *Extents* that do not include any part of the Eastern region. Variations per zone increase with the size of the *West CBD*, due to the larger number of new jobs. Peculiarities (*E40* and *E45* for *Small* and *Large West CBD* cases) will be discussed hereafter.

Real estate prices – Prices of non-residential buildings depend on population and job densities. The spatial structure of their variation between *Extents* is similar to that observed for jobs and therefore important, though with a larger level of noise due to the observed random pattern of households. Variations at both regional (Table 5) and zonal (Figure 6) levels show in most cases positive differences close to the CBD, meaning that an increase in real estate prices is lower in these zones for *Smaller Extents* than for the *Complete* Extent.

Table 5. Final values per region for the *Complete Extent* and relative differences (%) with *Small Extents* ($\alpha \leq$ 0.05*, 0.01**, and 0.001***).

		Equal-size CBDs			Small West CBD			Large West CBD		
Indicator	Extent	West CBD	Suburban	East CBD	West CBD	Suburban	East CBD	West CBD	Suburban	East CBD
Households	60	0.46	0.08	0.046	0.31	0.08	0.61	0.61	0.08	0.31
(%)	50	*0.002*	*−0.03**	*0.002*	*−0.01***	*−0.04**	*0.02****	*0.001*	*0.03***	*−0.02****
	45	*−0.001*	*−0.007*	*0.01*	*−0.005*	*−0.03*	*0.02****	*0.003*	*−0.0001*	*−0.01*
	40	*0.002*	*0.001*	*−0.02*	*0.005*	*−0.02*	*0.002*	*0.001*	*−0.005*	*−0.01*
	35	*0.006***	*−0.04***		*0.01****	*−0.06****		*0.01****	*−0.07****	
	30	*0.003**	*−0.04**		*−0.005**	*0.05**		*0.01****	*−0.1****	
	25	*0*			*0*			*0*		
Jobs	60	0.50	0	0.50	0.35	0	0.65	0.65	0	0.35
(%)	50	*−2.11****	*0*	*2.27****	*−12.7****	*0*	*7.3****	*2.96****	*0*	*−5.9****
	45	*0.77****	*0*	*−11.1****	*−0.02*	*0*	*0.17*	*0.24****	*0*	*−6.5****
	40	*0.05****	*0*	*−59.6****	*−0.001****	*0*	*0.77****	*−0.004***	*0*	*10.7***
	35	*0*	*0*		*0*	*0*		*0.001*	*0*	
	30	*0*	*0*		*0*	*0*		*0*	*0*	
	25	*0*			*0*			*0*		
Price	60	24 757	4 499	24 760	16 698	4 541	33 209	33 211	4 539	16 705
(Euros)	50	*−1.1****	*−0.01*	*−1.17****	*−5.25****	*−0.24*	*−5.68****	*−2.65****	*−0.05*	*−2.75****
	45	*−0.46****	*0.38*	*−0.24****	*11.9****	*−0.1*	*8.1****	*3.9****	*0.11*	*2.54****
	40	*−0.43****	*0.34**	*0.36****	*8.39****	*0.31**	*1.53****	*2.45*	*−0.06*	*0.17*
	35	*−0.64****	*0.55*		*−0.70****	*−0.62****		*−1.7****	*−0.14*	
	30	*−1.1****	*0.13*		*−1.1****	*−1.1****		*−2.4****	*−0.52**	
	25	*−2.1****			*−1.9****			*−3.8****		
Commuting	60	46.9	64.1	45.4	35.5	79.9	77.7	62.5	68.2	34.6
Times	50	*−6.7***	*−16.1****	*−30.9****	*−7.0****	*−35.6****	*−49.0****	*−10.1**	*−8.9****	*−15.7***
(Minutes)	45	*−7.4***	*−7.4****	*−7.6***	*−17.4****	*−47.3****	*−54.4****	*−8,7*	*12.9****	*52.3****
	40	*−9.2****	*−3.3****	*−18.9****	*−18.5****	*−40.0****	*−34.1****	*−9.0**	*15.1****	*75.1****
	35	*−12.8****	*−4.7****		*−21.1****	*−41.9****		*−9.1**	*14.5****	
	30	*−16.3****	*−11.9*		*−23.0****	*−39.5****		*−13.5***	*−3.4****	
	25	*−20.7****			*−26.6****			*−16.6****		

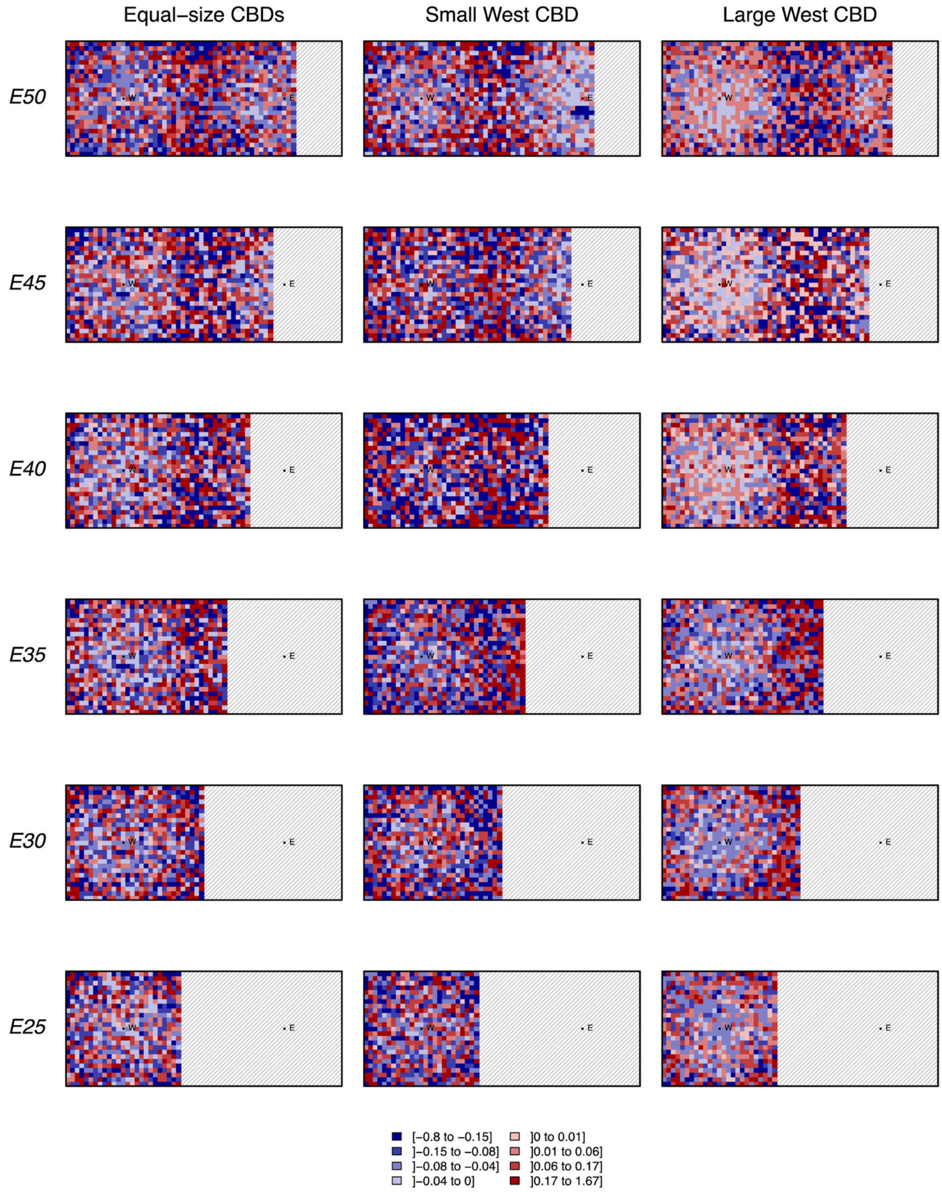

Figure 4. Number of households (value: (Complete − Exx)/Complete in t_{10}, with Exx = E25 to E50, in %; discretization: quantiles).

The parameter estimates of the real estate price sub-model vary from one case study to another, therefore, affecting the evolution of these prices. In particular, the parameter estimates of the population density (log) are slightly larger for *E40* and *E45*, which explain the peculiarities observed for these *Extents* for real estate prices and jobs.

Commuting travel time – The main transport output from the model, for it is simulated after the land-use part, is even more influenced by differences in extents. Observed variations are greater than for the previous indicators and their spatial structure is clear[3] and

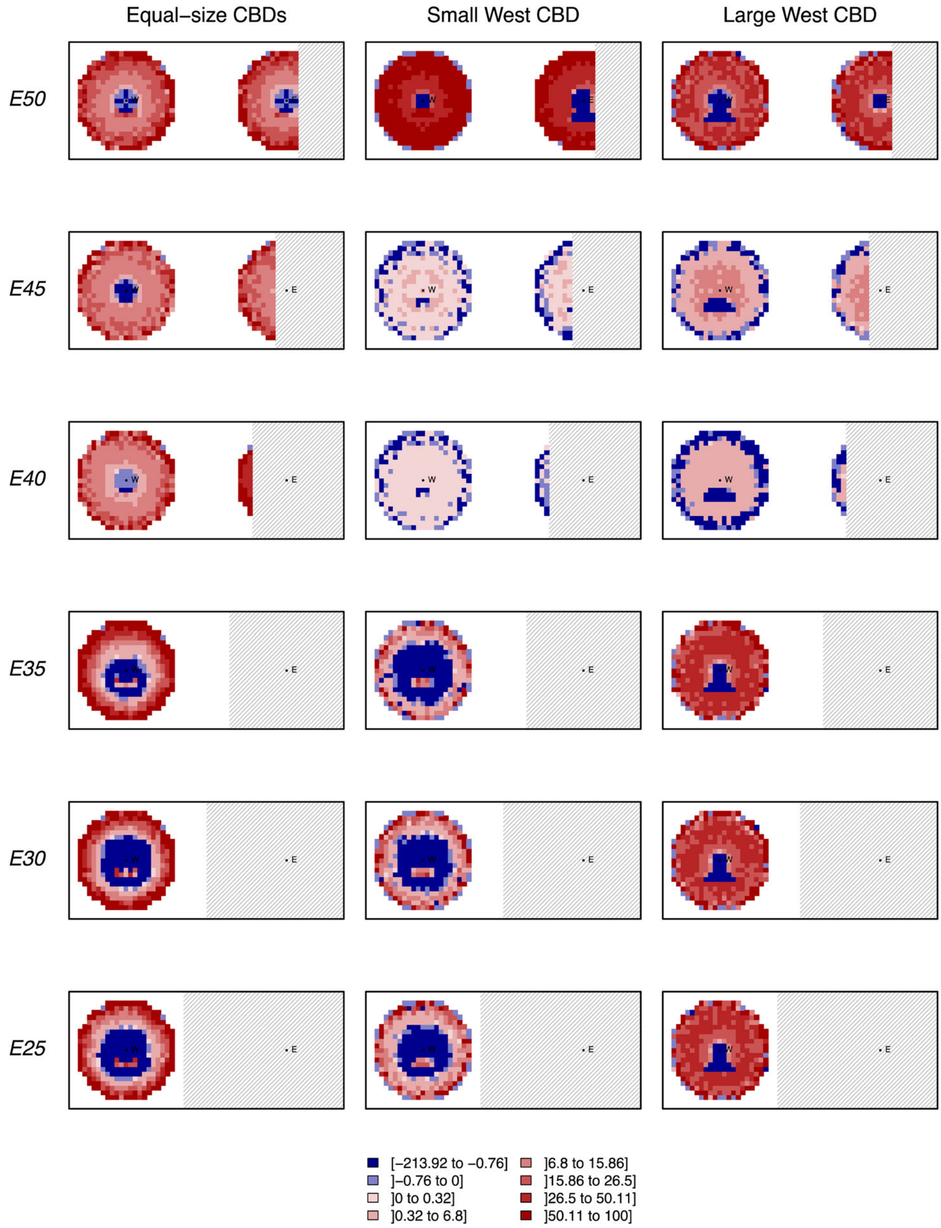

Figure 5. Number of jobs (value: (Complete – Exx)/Complete in t_{10}, with Exx = E25 to E50, in %; discretization: quantiles: White area: no jobs in t_0).

highly influenced by the relative size of the *West CBD*. The commuting time per region increases close to the largest CBD, and decreases in the zones furthest from this CBD (Table 5). Hence, for the *Small west CBD* case, a large increase in commuting time (negative differences with the *Complete* extent) is observed for all zones located East of this CBD, while for the case of the *Large west CBD* a decrease is observed (Figure 7). These variations are larger for *Extents* E35 to E45 that include a portion of the eastern region, but not the *East CBD* itself (Table 5).

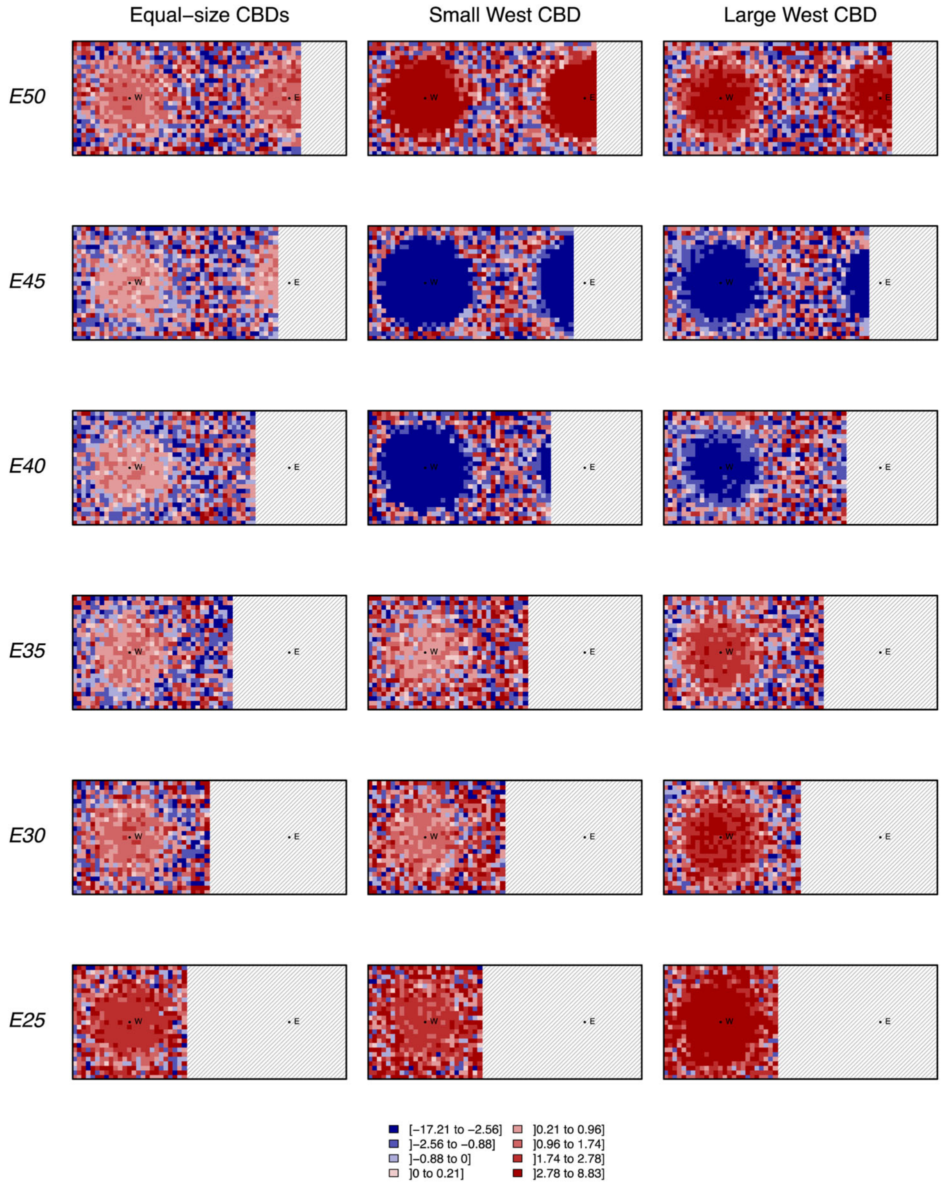

Figure 6. (non-residential) real estate prices (value: (Complete − Exx)/Complete, with Exx = E25 to E50, evolution between t_0 and t_{10} in %; discretization: quantiles).

3.4. Implications

The different *Extents* of our synthetic geography were not designed with a concern for realism beyond the Alonso–Muth–Mills standard trade-off, but to provide a systematic evaluation of the influence of the delineation of cities. Interestingly, we end up with quite realistic situations such as *Extents E30* and *E35* that consist of adding to the studied CBD a rural area with few or no functional links (commuting) with the CBD. Such a situation is found in several of the papers reviewed in Section 2 (e.g. the case of

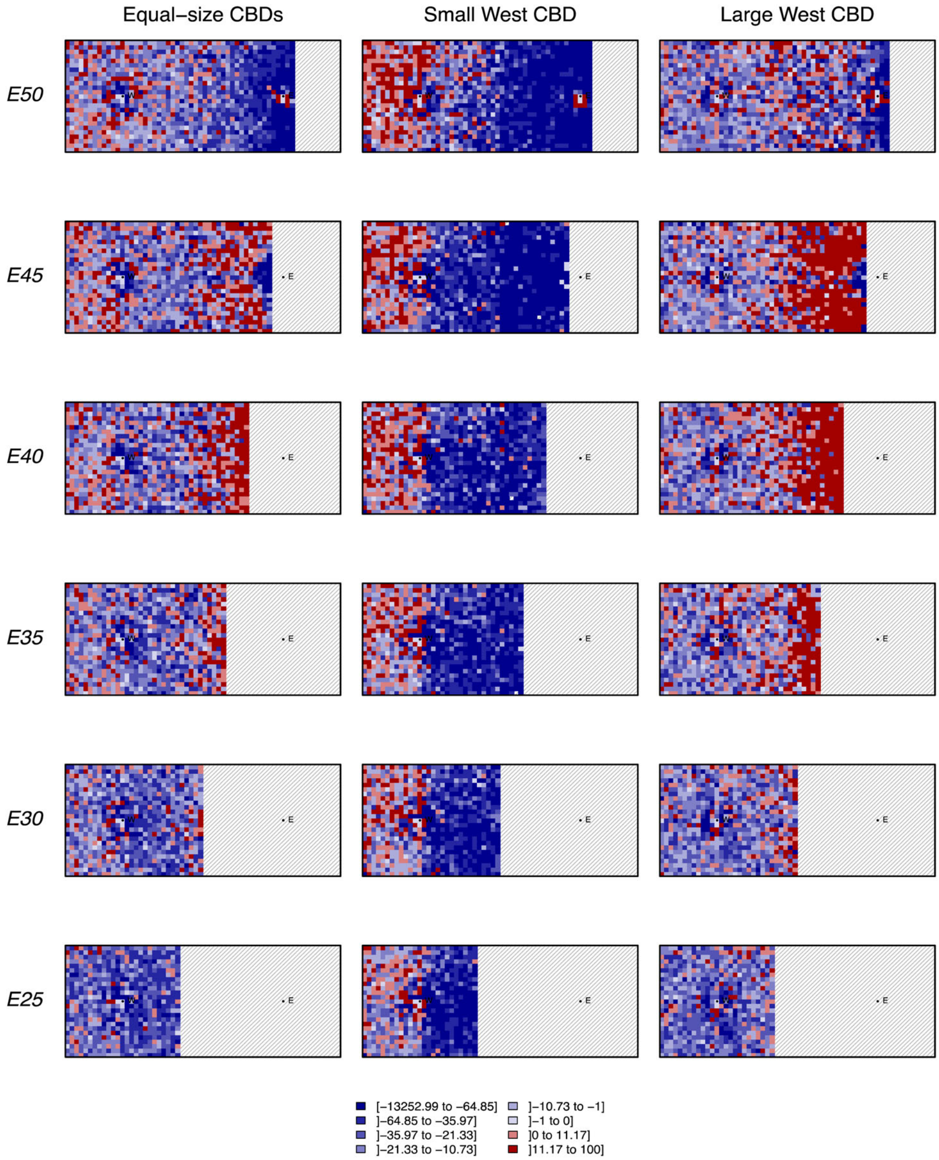

Figure 7. Home-to-work commuting time (value: (Complete − Exx)/Complete in t_{10}, with Exx = E25 to E50, in %; discretization: quantiles).

Paris). On the opposite, *Extents E40* to *E50* include a portion of another CBD within the studied urban area (see, for instance, the case of Brussels mentioned in Section 2).

Our simulations suggest that the inclusion of rural areas has only a limited influence on model outcomes (*Extent E30*, where the relative variations can be associated with noise). Nevertheless, we have demonstrated that if the study area is not well associated with a functional region, it dramatically impacts the location of agents, especially jobs, and subsequent model outputs, in particular, commuting time. More precisely, a bias arises (1) when the study area fails to encompass some portions of the area of influence of one

CBD (*Extent E50*, where peripheral areas of the East CBD are excluded), and (2) when the study area includes some portions of the catchment area of a nearby CBD but excludes its centre (*Extents E35* to *E45*, do not include the East CBD itself). These cases show major discrepancies to the *Complete Extent* that can lead to a wrong understanding of transport and land-use coefficients and outputs. Moreover, from the *Extent E50*, we see that the magnitude of the bias is proportional to the size of the excluded CBD (the bias increases when the East CBD is larger than the one under focus), pointing out clearly the importance of accurately delimiting a functional area within its wider geographical environment and nearby competing cities.

4. Discussion and recommendations

In the scientific literature related to LUTI applications (see Section 2), the size and number of spatial units is sometimes described, but very few authors discuss the reason for choosing a spatial extent for their urban case study. This is a remarkable vet worrying result given the popularity of LUTI models in geography and urban studies. Most likely, the political/administrative authority that supports the development and application of an LUTI model will require the area to encompass only its activity area, whatever the functional reality of the delineation. Similarly, the land-use and transport data needed for an LUTI application can be directly associated with a particular regional setting, independent of any functional reasoning. Beyond these very pragmatic reasons, there seems to be a presumption in research articles that the delineation of a study area is not very problematic. A probable reason for this lack of precision and interest may be the fact that the major ingredients of the urban system (number of jobs, of residents, and the transport infrastructure) are included in the model. Another believe could be that because densities decrease with distance, cutting the more distant locations out of the model has only benign effects.

Section 3 shows that these assumptions cannot hold even with a simple urban structure. The delineation of the study area plays a capital role in the results and the consequences for not controlling the city delineation are clearly underestimated in the literature. This being said, defining what is a coherent functional urban system is still far from obvious. There is a large amount of literature and a wide variety of existing methods for delineating a city (see Dujardin et al., 2007; Thomas et al., 2013; or Jones et al., 2015a for review and examples). In light of our findings and keeping in mind, the practical objectives of most LUTI models, we can make three main recommendations: (1) modellers should search for an optimal delineation of the study area based on commuting flows prior to any simulation; (2) as a corollary, a clear coupling of the model components must then be made with the “Rest of the World”; and (3) a detailed reporting of aggregates (population, jobs, etc.) and geographical pattern of densities, prices and flows should be communicated for transparency and understanding.

An optimal endogenous delineation – Given the still very dominant role of the trade-off between commuting costs and real estate prices, commuting flows should be the base for delineating the application area of an LUTI model. As long as jobs’ places are known, this criterion can be used independently of whether the city is mono or polycentric. Typically, a delineation method could be based upon the partition of a wider origin – destination matrix for commuting, and use an endogenous criterion to produce clusters with maximal intra-group flows and minimal inter-group flows. Examples can be found in

Farmer and Fotheringham (2011), Coombes (2014), and Thomas et al. (2013). Obviously the level of spatial detail then comes into play as well. Ideally, the base matrix should be as detailed as possible and the robustness of the partition to aggregation controlled for.

Defining an optimal delineation by such method meets three desirable characteristics for LUTI models. First, commuting is the product of the spatial mismatch between residential locations and employment centres. The focus is therefore set on the interactions between places through the transport infrastructure, as assumed by the "transport and land-use feedback cycle" (Simmonds, Waddell, & Wegener, 2013). Second, commuting flows typically produce large delineations, which is consistent with the extended urban area favoured in most LUTI models' applications (see Section 2). Third, the proposed approach seeks to create a study area with strong internal ties while lowering links with the Rest of the World that are not simulated with the same detail, thus reducing bias from the border effect.

Coupling with the "Rest of the World" – In some LUTI models external boxes are used to represent outer zones, for which results are not under focus. Mostly, this is done in the transport system, e.g. with in/out commuting flows or exogenous background traffic. Such interactions can be found in MEPLAN, IRPUD, or DELTA LUTI models (see Bosredon, Doson, & Simmonds, 2009; Echenique et al., 1990; Simmonds et al., 2013; Wagner & Wegener, 2007).

Defining what is outside of a modelling system is a corollary to defining what is inside. In facts, the same method of origin – destination matrix partition suggested for the delineation of the study area can be used to draw external zones. How many of them are needed is then again a question of significance of flows with the modelled system and internal consistency of the available "Rest of the World". The more spatial disaggregation there will be of the external system, the more scenarios related to a wider territory (e.g. as in Eradus et al., 2002; Zondag et al., 2015) can be tested, though mainly as input. However, the question will then be raised on where to put the limit again and what the benefits are of having two levels of details for a single LUTI model: one disaggregated and full feedback effects and one more aggregated at the outskirts with fewer spatial details and less interactions.

Given the difficulty of calibrating and reaching robustness for a single precise system and known effects of spatial aggregation (see Jones, 2016), our opinion is rather to keep the external world as simple as possible as an attractor/sender of flows. The purpose of LUTI models is to handle land-use and transport feedbacks in a dynamic fashion for prospective research. If one wants to question more regional interactions' effects, relying on more standard spatial interaction models (even in a static manner) seems rather sufficient and typically the detail of traffic assignment, and the precise positioning of households and prices are unnecessary burdens. Moreover, interpretation of a single coherent urban region is easier to relate to processes in urban economics.

Transparency – Reproducibility of LUTI applications is often deemed insufficient due to the lack of details on the methodological framework (te Brömmelstroët, Pelzer, & Geertman, 2014; Lee, 1973; Nguyen-Luong, 2008; Saujot et al., 2016). Our results demonstrate that this criticism is also valid for the choice of geographical delineations. Relying on a clearly stated methodology for optimising the definition of a study area would allow for better liaising findings across case studies. Especially it would help showing similarities and dissimilarities for particular sets of regions – e.g. large or small city regions, mono-or poly-centric – while being certain that the geographical delineation has not impacted the comparison. Furthermore, regulation and planning can have decisive impacts on urban and transport development. Understanding the impact of specific policies is only possible after a good knowledge of the functioning of the study area is obtained, which in turn requires transparent and accurate understanding of its geographical context. Finally, with clearly stated methods for optimising the definition of study areas, results of LUTI applications can then be discussed in light of other research fields. Typically knowledge transfer would be easier with other empirical and theoretical research on urban expansion forms or in the field of urban and transport economics. While offering more comparative power, coming-up with an explicit and optimised delineation methodology could progress the theoretical urban agenda.

5. Conclusion

We have shown that there is a lack of care in the European LUTI research community for properly delineating case study applications. This is a problem that not only challenges the reproducibility and comparability of results but also weakens the results themselves. The call made here for a more careful delineation of study areas is not only of a theoretical interest (making LUTI models capable of contributing to generalisable scientific knowledge), but also has practical implications. For example, a remarkable result in many applications of LUTI models to European cities is their strong inertia (e.g. Cabrita et al., 2015; Nguyen-Luong, 2008): land-use adjustments to external (including policy) perturbation have a very long-time span (see Wegener, Gand, & Vannahme, 1986). Incorrect delineations of study areas may *de facto* contribute to such stability, because areas that are not under the influence of the city under focus and strongly impacted by policy scenarios, are included in the process.

We definitely need LUTI models for understanding, planning, and anticipating transportation and urban development problems within city regions. LUTI models, require numerous and diverse competences that would gain to be better integrated as the field is now in a maturing phase. Very recently *Moebius* (Luxembourg; see Lord, Frémond, Biltgen, & Gerber, 2015) and *Sustaincity* (Paris, Zürich, and Brussels; see Bierlaire, de Palma, Hurtubia, & Waddell, 2015) research projects illustrated the difficulties in applying LUTI models in Europe. There are of course many other technical and conceptual challenges than the geographical aspects assessed here. However, the delineation of functional urban areas, a major concern in geographical science for decades, has been too much ignored in European LUTI applications so far. We show this is detrimental to model results and comparability. Hopefully, there is space for improvements and the guidelines we offer constitute only a very first step towards “better spatial practices” in LUTI models.

Notes

1. We use the zone-version of the *UrbanSim* model. For consistency purpose, the cells of the grid will be referred to as “zones”.
2. Since all zones have the same size, absolute numbers are equivalent to densities.
3. Note that some noise appears in the estimates of commuting times, since *MATsim* uses only 25% of the agents (the maximum compatible with the computer power required, which is still larger than the 10% recommended by Nicolai & Nagel, 2015).

Acknowledgements

The authors are grateful to the funding agencies that made this research possible. We also thank two anonymous referees for their useful comments on a previous version of this work, as well as Kerry Schiel (University of Luxembourg) for proof reading the manuscript. Isabelle Thomas is grateful to LISER for the having offered the opportunity of a research stay, during which this work was initiated.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study is both part of the MOEBIUS 2010–2014 research project (MObilities, Environment, Behaviours, Integrated in Urban Simulation) co-funded by the National Research Fund of Luxembourg (FNR CO09/SR/07) and the Luxembourg Institute for Socio-Economic Research (LISER), and of the SUSTAINCITY 2010-2013 research project (Micro-simulation for the prospective of sustainable cities in Europe) funded by the European Union (FP7-SSH 244557).

ORCID

Geoffrey Caruso http://orcid.org/0000-0001-9538-2328

References

Abdel-Rahman, H. M., & Anas, A. (2004). Theories of systems of cities. In J. V. Henderson, & J. F. Thisse (Eds.), *Handbook of regional and urban economics. Volume 4 cities and geography* (pp. 2293–2340). North-Holland: Elsevier.

Acheampong, R. A., & Silva, E. A. (2015). Land use-transpirt interaction modeling: A review of the literature and future research directions. *Journal of Transport and Land Use*, *8*(3), 11–38.

Alonso. (1964). *Location and land use*. Cambridge, MA: Harvard University Press.

*Anas, A. (2013a). A summary of the applications to date of RELU-TRAN, a microeconomic urban computable general equilibrium model. *Environment and Planning B: Planning and Design*, *40* (6), 959–970.

Anas, A. (2013b). A response to the guest editorial: Economics as the science for urban modelling. *Environment and Planning B: Planning and Design*, *40*, 955–958.

*Anderstig, C., & Mattsson, L.-G. (1992). Appraising large-scale investments in a metropolitan transportation system. *Transportation*, *19*(3), 267–283.

Arauzo-Carod, J. M., & Manjón-Antolín, M. (2004). Firm size and geographical aggregation: An empirical appraisal in industrial location. *Small Business Economics*, *22*(3-4), 299–312.

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

Batty, M. (2005). *Cities and complexity: Understanding cities with cellular automata, agent-based models, and fractals*. Cambridge, MA: MIT Press.

Batty, M. (2008). The size, scale, and shape of cities. *Science*, *319*, 769–771.

*Batty, M., Vargas, C., Smith, D., Serras, J., Reades, J., & Johansson, A. (2013). SIMU-LACRA: Fast land-use-transportation models for the rapid assessment of urban futures. *Environment and Planning Part*, *40B*(6), 987–1002.

Berry, B. J., & Lamb, R. F. (1974). The delineation of urban spheres of influence: Evaluation of an interaction model. *Regional Studies*, *8*(2), 185–190.

Bertaud, A., & Malpezzi, S. (2003). *The spatial distribution of population in 48 world cities: Implications for economies in transition*. Center for Urban Land Economics Research, University of Wisconsin.

Retrieved from http://www2.lawrence.edu/fast/finklerm/Complete%20Spatial%20Distribution%20of%20Population%20in%2050%20World%20Ci.pdf

Bierlaire, M., de Palma, A., Hurtubia, R., & Waddell, P. (Eds.). (2015). *Integrated transport & land use modelling for sustainable cities*. Lausanne: EPFL Press.

*Bonin, O., & Tomasoni, L. (2015). Evaluation of a transit-oriented development scenario in a medium-sized French city by simulation models. *International Journal of Transportation*, *3*(1), 91–112.

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.

Brueckner, J. K. (2001). Urban sprawl: Lessons from urban economics. *Brookings-Wharton Papers on Urban Affairs*, *2001*(1), 65–97.

Brueckner, J. K., Thisse, J. F., & Zenou, Y. (1999). Why is central Paris rich and downtown Detroit poor? An amenity-based theory. *European Economic Review*, *43*(1), 91–107.

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.

*Burgos, J. (1994). Integrated land-use and transport models in the basque country. *Environment and Planning B: Planning and Design*, *21*(5), 603–610.

Cabrita, I., Gayda, S., Hurtubia, R., Efthymiou, D., Thomas, I., Peeters, D., … Roder, D. (2015). Integrated land use and transport microsimulation for Brussels. In M. Bierlaire, A. de Palma, R. Hurtubia, & P. Waddell (Eds.), *Integrated transport and land use modelling for sustainable cities* (pp. 373–412). Lausanne: Routledge/EPFL Press.

Caruso, G., Peeters, D., Cavailhès, J., & Rounsevell, M. (2007). Spatial configurations in a Periurban city. A cellular automata based micro economic model. *Regional Science and Urban Economics*, *37*, 542–567.

Caruso, G., Vuidel, G., Cavailhès, J., Frankhauser, P., Peeters, D., & Thomas, I. (2011). Morphological similarities between DBM and a microeconomic model of sprawl. *Journal of Geographical Systems*, *13*(1), 31–48.

Cervero, R. (2002). Built environments and mode choice: Towards a normative framework. *Transportation Research Part D: Transport and Environment*, *7*(4), 265–284.

Chakraborty, A., Wilson, B., & bin Kashem, S. (2015). The pitfalls of regional delineations in land use modeling: Implications for Mumbai region and its planners. *Cities*, *45*, 91–103.

Cheshire, P. C., & Gornostaeve, G. (2002). Villes et regions urbaines: Des comparaisons fiables doivent reposer sur des territories comparables. *Les Cahiers de l'IAURIF*, *135*, 13–32.

Coombes, M. (2014). From city-region concept to boundaries for governance: The English case. *Urban Studies*, *51*(11), 2426–2443.

*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*(3), 153–165.

Cörvers, F., Hensen, M., & Bongaerts, D. (2009). Delimitation and coherence of functional and administrative regions. *Regional Studies*, *43*(1), 19–31.

De Borger, B., Proost, S., & Van Dender, K. (2008). Private port pricing and public investment in port and hinterland capacity. *Journal of Transport Economics and Policy (JTEP)*, *42*(3), 527–561.

Derycke, P. H., Huriot, J. M., & Pumain, D. (Eds.). (1996). *Penser la ville. Théorie et modèles*. Paris: Anthropos.

*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*(2), 85–105.

Dujardin, C., Thomas, I., & Tulkens, H. (2007). Quelles frontières pour Bruxelles? Une mise à jour. *Reflets et perspectives de la vie économique*, *46*(2–3), 155–176.

*Echenique, M., Flowerdew, A., Hunt, J., Mayo, T., Skidmore, I., & Simmonds, D. (1990). The MEPLAN models of bilbao, leeds and dortmund. *Transport Reviews*, *10*(4), 309–322.

*Eradus, P., Shoemakers, A., & van der Hoorn, T. (2002). Four applications of the TIGRIS model in the netherlands. *Journal of Transport Geography*, *10*(2), 111–121.

Farmer, C., & Fotheringham, A. S. (2011). Network-based functional regions. *Environment and Planning A*, *43*, 2723–2741.

Federal Register. (2000). Standards for defining metropolitan and micropolitian statistical areas. *Federal Register* 65.

Federal Register. (2002). Qualifying urban areas for Census 2000: Notices. *Federal Register* 67.

Fotheringham, A. S., & Wong, D. W. (1991). The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, *23*(7), 1025–1044.

Griffith, D. A. (1983). The boundary value problem in spatial statistical analysis. *Journal of Regional Science*, *23*(3), 377–387.

*Guzman, L., de la Hoz, D., & Circella, G. (2015). Evaluations of synergies from transportation policy packages using a social welfare maximization approach: A case study for Madrid, Spain. *Case Studies on Transport Policy*, *3*, 99–110.

Hall, P. (2007). Delineating urban territories. Is this a relevant issue? *Cities and Networks in Europe: A Critical Approach of Polycentrism, Paris: John Libbey Eurotext*, 3–14.

Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: Views from urban planning. *American Journal of Preventive Medicine*, *23*(2), 64–73.

Hohenberg, P. M., & Lees, L. H. (1986). *The making of urban Europe 1000—1950*. Cambridge, MA: Harvard University Press.

*Hunt, J. (1994). Calibrating the Naples land-use and transport model. *Environment and Planning B: Planning and Design*, *21*(5), 569–590.

Hunt, J. D., Miller, E. J., & Kriger, D. S. (2005). Current operational urban land-use transport modelling frameworks. *Transport Reviews*, *25*(3), 329–376.

Hunt, J. D., & Simmonds, D. C. (1993). Theory and application of an integrated land use and transport modelling framework. *Environment and Planning B: Planning and Design*, *20*(2), 221–244.

Jones, J., Peeters, D., & Thomas, I. (2015a). Is cities delineations a prerequisite for urban modelling? The example of land price determinants in Brussels. *Cybergeo: Revue Européenne de Géographie*, Retrieved from https://cybergeo.revues.org/26899

Jones, J., Thomas, I., & Peeters, D. (2015b). Forecasting employment location choices by discrete choice models: A sensitivity analysis to scale and implications for LUTI models. *REGION*, *2*(1), 67–93.

Jones, J. (2016). *Spatial bias in LUTI models* (Doctoral thesis). Presse Universitaire de Louvain. 283p.

Kwan, M. P. (2012). The uncertain geographic context problem. *Annals of the Association of American Geographers*, *102*(5), 958–968.

Lee, Jr D. B. (1973). Requiem for large-scale models. *Journal of the American Institute of Planners*, *39*(3), 163–178.

Lord, S., Frémond, M., Biltgen, R., & Gerber, P. (2015). Urban growth and land management in a context of sustainable development: Questioning the planning policies and scenarios beyond the rules. *Planning Theory and Practice*, *16*(3), 385–406.

Lowry, I. (1964). *A model of metropolis* (Technical Report RM-4035-RC). The RAND Corporation, California: Santa Monica.

*Mackett, R. (1990). The systematic application of lilt model to Dortmund, Leeds and Tokyo. *Transport Reviews*, *10*(4), 323–338.

*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: Journal of the Transportation Research Board*, *1924*, 129–138.

Nagel, K., Grether, D., Beuck, U., Chen, Y., Riser, M., & K, A. (2008). Multi-agent transport simulations and economic evaluation. *Journal of Economic and Statistics*, *228*(2+3), 173–194.

Nguyen-Luong, D. (2008). *An integrated land use-transport model for the Paris region (SIMAURIF): Ten lessons learned after four years of development*. Retrieved from http://web.mit.edu/11.521/proj08/readings/D_Mes_documentsDNLpredit3ERSA_2008article_SIMAURIF_10_lessons.pdf

Nicolai, T., & Nagel, K. (2015). Integration of agent-based transport and land use models. In M. Bierlaire, A. de Palma, R. Hurtubia, & P. Waddell (Eds.), *Integrated transport & land use modelling for sustainable cities* (pp. 333–353). Lausanne: EPFL Press.

Noth, M., Borning, A., & Waddell, P. (2003). An extensible modular architecture for simulating urban development, transportation and environmental impact. *Computer, Environment and Urban Systems*, *27*, 181–203.

NUREC – Network on Urban Research in the European Union. (1994). *Atlas of agglomerations in the European Union, Volumes I–III*. Duisburg: NUREC Edition.

Openshaw, S. (1984). Ecological fallacies and the analysis of areal census data. *Environment and Planning A*, *16*(1), 17–31.

de Palma, A., Picard, N., & Motamedi, K. (2015). Application of UrbanSim in Paris (Ile-de-France) case study. In M. Bierlaire, A. de Palma, R. Hurtu- bia, & P. Waddell (Eds.), *Integrated transport and land use modelling for sustainable cities* (pp. 413–460). Lausanne: Routledge/EPFL Press.

Parr, J. B. (2007). Spatial definition of the city: Four perspectives. *Urban Studies*, *44*(2), 381–392.

*Patterson, Z., & Bierlaire, M. (2010). Development of prototypes UrbanSim models. *Environment and Planning B: Planning and Design*, *37*(2), 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*(2), 5–37.

Paulley, N. J., & Webster, F. V. (1991). Overview of an international study to compare models and evaluate land use and transport policies. *Transport Reviews*, *11*(3), 197–222.

*Pozoukidou, G. (2014). Land use and transport interaction models: Application perspectives for the city of Thessaloniki. *Spatium International Review*, *32*, 7–14.

Pumain, D. (Eds.). (2006). *Hierarchy in natural and social sciences*. Dordrecht: Springer.

Raciti, S. M., Hutyra, L. R., Rao, P., & Finzi, A. C. (2012). Inconsistent definitions of “urban” result in different conclusions about the size of urban carbon and nitrogen stocks. *Ecological Applications*, *22*(3), 1015–1035.

Saujot, M., de Lapparent, M., Arnaud, E., & Prados, E. (2016). Making land use–transport models operational tools for planning: From a top-down to an end-user approach. *Transport Policy*, *49*, 20–29.

Schirmer, P., Zollig Renner, C., Mülcer, K., & Axhausen, K. (2015). Land use and transport microsimulation in the canton of Zurich using UrbanSim. In M. Bierlaire, A. de Palma, R. Hurtubia, & P. Waddell (Eds.), *Integrated transport and land use modelling for sustainable cities* (pp. 485–536). Lausanne: Routledge/EPFL Press.

Simmonds, D., Waddell, P., & Wegener, M. (2013). Equilibrium versus dynamics in urban modeling. *Environment and Planning B: Planning and Design*, *40*, 1051–1070.

Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. (2005). *Thematic cartography and geographic visualization* (2nd ed). Upper Saddle River, NJ: Prentice-Hall Series in Geographic Information Science.

Tannier, C., & Thomas, I. (2013). Defining and characterising urban boundaries: A fractal analysis of theoretical and real Belgian cities. *Computers, Environment and Urban Systems*, *41*, 234–248.

Thomas, I., Cotteels, C., Jones, J., & Peeters, D. (2013). “Revisiting the Extension of the Brussels Urban Agglomeration: New Methods, New Data … New Results?” *Belgeo* [online] 1–2. Retrieved from: http://belgeo.revues.org/6074.

*Vold, A. (2005). Optimal land use transport planning for the greater Oslo area. *Transportation Research Part A*, *39*(6), 548–565.

Waddell, P. (2000). A behavioural 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. (2011). Integrated land use and transportation planning and modelling: Addressing challenges in research and practice. *Transport Reviews*, *31*(2), 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*(1), 43–67.

*Wagner, P., & Wegener, M. (2007). Urban land use, transport and environment models: Experience with an integrated microscopic approach. *disP*, *170*(3), 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.

Webster, F. V., & Dasgupta, M. (1991). Land Use and Transport Interactions: Report of the ISGLUTI Study. *CONTRACTOR REPORT, 295, ISSN: 0266-7045;*, (295).

Wegener, M. (2011). From macro to micro: How much micro is too much? *Transport Reviews*, *31*(2), 161–177.
Wegener, M., Gand, F., & Vannahme, M. (1986). The time scale of urban change. In B. Hutchinson, & M. Batty (Eds.), *Advance in urban system modeling* (pp. 175–197). North-Holland: Amsterdam.
*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.
*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.

Note: the papers reviewed in the meta-analysis are preceded by an asterisk (*).

Appendix 1 – Initial structure of the synthetic urban areas

The synthetic city is developed on a featureless landscape. We decided to define the utility of households and jobs as in the Alonso–Muth–Mills model, i.e. increasing with the accessibility to jobs and decreasing with real estate prices. However, in *UrbanSim*, this utility level derives from the independent factors selected in the household and employment location sub-models. Therefore, we set the initial distribution of households and jobs per zone using a function of the distance to the CBD. The functional form is obtained by trial and error, until the parameter estimates have the desired signs within the real estate model of *UrbanSim*. The mathematical forms obtained are an inverted logistic curve for households, and a negative exponential curve for non-home-based jobs. An advantage of this choice is that jobs are more highly concentrated towards the centre than households, which is consistent with the multi sector version of the Alonso–Muth–Mills model.

Let H be the total number of households h, in t_0 and J is the total number of jobs j. It is assumed that every household includes two workers, hence $J = 2H$. Equations (A1), for households, and (A2), for non-home-based jobs, allow for determining the attraction potential of a zone (cell) i, denoted by P_i. Hereafter, α denotes the relative size of the CBDs, β is the distance-decay parameter (unitary in this case), d_{in} the distance between i and the CBD n, and N is the number of CBDs.

$$P_i(h) = \left(\sum_{n=1:N} \alpha \frac{1}{1 + e^{-\beta d_{in}}} \right) / N, \tag{A1}$$

$$P_i(j) = \left(\sum_{n=1:N} \alpha e^{-\beta d_{in}} \right) / N. \tag{A2}$$

The potential P_i(h) and P_i(j) are re-scaled between 0 and 1. The total number of households h_i and non-home-based jobs j_i in a zone i in t_0 are then equal to HP_i(h) and $0.95JP_i$(j) (among all jobs 95% are non-home-based jobs). Home-based jobs are distributed between zones proportionally to the household' density. To simulate the bi-centric settings, we set $N = 2$ and $\alpha = 2$, 1 or 0.5.

In zone versions of *UrbanSim*, agents are located in buildings, which in turn are situated in a zone. In our experiment, buildings are mono-functional (i.e. purely residential or non-residential) and limited to *Houses* and *Offices*. Their characteristics depend on the number of agents per zone. For residential buildings (*Houses*), the number of existing residential units (i.e. dwellings) in t_0 is given by $ru_i = h_i(1+vr)$ with vr the average, long-term, vacancy rate (set here to 10%). The residential units capacity ruc_i is equal, for all *Houses*, to the maximal value of ru_i. Hence, we assume that the residential developments' capacity is null in the CBD and increase with the Euclidean distance to the CBD. Finally, the average value per residential unit, $p_i(ru)$ is defined by Equation (A3), where $\mu(h)$ denotes the average number of households per zone and C_{ru} is a constant (set to 50,000 € here). Note that its actual value has no practical importance: since the income level is uniform among households, only the spatial variations of the residential prices matter in their location choices.

$$p_i(ru) = C_{ru}\frac{h_i}{\mu(h)}€. \tag{A3}$$

Non-residential buildings (*Offices*) are characterized by their existing floor space for jobs, their floor space capacity (existing + developable) and their average price. This latter characteristic is the value of one-square metre of floor space, not the price of the entire building. This is a hard-coded assumption in the source code of *UrbanSim*. The non-residential surface in t_0, noted nr_i, is equal to $20j_i(1 + vr)$. In other words, we assume that each job requires a surface of 20 square metres. The non-residential surface capacity (nr_c) is equal to the maximal value of nr_i. The real estate prices in t_0 are given by Equation (A4), where $\mu(j)$ denotes the average number of jobs per zone, and C_{nr} is a constant equal to 100 €/m^2.

$$p_i(nr) = C_{nr}\left(\frac{h_i}{\mu(h)} + \frac{j_i}{\mu(j)}\right)€/\text{m}^2. \tag{A4}$$

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, do not include any residential units.

Two macro-economic parameters have to be defined in any application of *UrbanSim*. (1) The population growth is assumed to be linear, meaning that H_t, the number of households in t, is equal to $H(1 + g)t - 1$ with t being the number of year and g is the population growth rate. The control totals for jobs are derived from these values for households, following the rule that $J = 2H$. (2) The relocation rates are identical for households and jobs and set to 10% per year. Finally, a bi-directional road network connects the centroid of each zone to the centroid of all adjacent zones, on a von-Neumann neighbourhood. The length of each link is the Euclidean distance between the two centroids. Maximal speed is set to 13.88 m/s (i.e. 50 km/h, the maximal authorised speed in urban areas in the most European countries), and the capacity of each lane (one in every direction) to 500 vehicles per hour.