Integrated land use and transportation modelling and planning: A South African journey

Louis Waldeck
Council for Scientific and Industrial Research
louis.waldeck@gmail.com

Jenny Holloway
Council for Scientific and Industrial Research
JHollowa@csir.co.za

Quintin van Heerden (corresponding author)
Council for Scientific and Industrial Research
QvHeerden@csir.co.za

Abstract: Confronted by poverty, income disparities and mounting demands for basic services such as clean water, sanitation and health care, urban planners in developing countries like South Africa, face daunting challenges. This paper explores the role of Integrated land use and transportation modelling in metropolitan planning processes aimed at improving the spatial efficiency of urban form and ensuring that public-sector investments in social and economic infrastructure contribute to economic growth and the reduction of persistent poverty and inequality. The value of such models is not in accurately predicting the future but in providing participants in the (often adversarial) planning process with a better understanding of cause and effect between different components of the urban system and in discovering common ground that could lead to compromise. This paper describes how an Urban Simulation Model was developed by adapting one of the leading microsimulation models (UrbanSim) originating from the developed world to South African conditions and how the requirements for microscopic data about the base year of a simulation were satisfied in a sparse data environment by introducing various typologies. A sample of results from three case studies in the cities of Tshwane, Ekurhuleni and Nelson Mandela Bay between 2013 and 2017 are then presented to illustrate how modelling supports the planning process by adding elements of rational analysis and hypothesis testing to the evaluation of proposed policies.

Keywords: Integrated land-use and transportation modelling, developing countries, UrbanSim, OpenTripPlanner

1   Introduction

Confronted by poverty, income disparities and mounting demands for basic services such as clean water, sanitation and health care, urban planners in developing countries like South Africa (SA) face daunting challenges (Cervero, 2013). Given that urban planning can be intensely controversial due to the involvement of many institutional and non-institutional stakeholders with divergent values and mandates (Waddell, 2011), the notion of evidence-based planning has emerged as a means of avoiding some of the conflict. Based on the premise that policy decisions can be improved by rational analysis and hypothesis testing, evidence can take the form of surveys, analyses of past observations or using mathematical models to simulate the likely future outcome of proposed policy scenarios.

Integrated Land Use and Transportation (ILUT) models represent a specific class of mathematical models that have over the last three decades contributed to policy formulation by providing a means of rational analysis and hypothesis testing in the evaluation of proposed policies. The models were initially focused exclusively on land use and transportation issues but are increasingly being used to assess potential economic, social, and environmental impacts in search of more sustainable cities. The relative success of a proposed policy scenario in achieving one or more objectives can be evaluated by comparing a few predefined indicators or measures of performance. A policy scenario is a unique combination of policy alternatives and assumptions, such as estimates of population and employment growth in the study area. Policy alternatives may include anything from growth boundaries to strategic decisions about transportation and other infrastructure systems. A comprehensive list of which policy alternatives might be considered to address various economic, social, and environmental issues was compiled by Wegener (2007) from a study of ILUT projects in Europe.

The value of the "evidence" obtained from such models is not derived from accurately predicting the future, but from providing participants in the hypothesis-testing process with a better understanding of interactions between different components of the urban system, which ranks among the most complex production systems ever built. Furthermore, it facilitates finding common ground between different interest groups that could lead to compromise.

In this paper we describe a decade long (2007 – 2017) journey of the development, behavioral and empirical validation (defined by Waddell, 2011) and application of a South African ILUT model in 3 metropolitan municipalities to support evidence-based decision making. The model developed by the CSIR will be referred to as UrbanSim-SA to accentuate the resemblance with UrbanSim-E; developed as part of the much bigger EU-funded SustainCity project (2010 – 2013) involving collaboration between 12 research institutions.

In Section 2 we describe material aspects of context and government policy in South Africa to provide non-South African readers with a better appreciation of the policy scenarios that UrbanSim-SA must be able to respond to. In Section 3 we consider related work from published applications of UrbanSim outside the USA, where it was developed. In Section 4 we describe how UrbanSim was adapted to South African conditions and reasons for linking it with a modified version of OpenTripPlanner as transport model. Arguably the biggest challenges to applying microsimulation models anywhere in the world, are the availability and effort required to assemble detailed data for the base year of a simulation at the level of individual parcels, buildings, households, etc. In Section 5 we describe how this was achieved in the sparse data environment characteristic of developing countries. Section 6 describes how both the model and the data preparation processes were empirically validated by simulating a period in the past (2001 to 2011) and comparing the results with actual growth patterns that occurred during the same period according to two reputable sources. In Section 7 we provide examples of the results obtained from simulation projects in the Cities of Tshwane, Ekurhuleni, and Nelson Mandela Bay between 2013 and 2017, each with a description of local context and some of the policy scenarios that were simulated.

2   Context

In addition to the daunting challenges posed by poverty, income disparities and demands for basic services, planning in South Africa is also heavily influenced by history, specifically the ideology of apartheid introduced by the National Party government when it came to power in 1948. Apartheid called for the separate development of different racial groups in South Africa. On paper it appeared to call for equal development and freedom of cultural expression but forced different racial groups to live separately and develop unequally (South African History Online, 2016). It may come as no surprise then that current planning practices are aimed squarely at resolving the inherited spatial inefficiencies in urban form.

It is generally accepted that the biggest inefficiency is that poor households live on the periphery of towns and cities. Cervero (2013) notes that spatial mismatches between where the needy live and where formal jobs with livable wages are located, occur in all cities but that they are more pronounced in most of the developing world.

In these cities, the poor live mainly on the fringes and effectively trade off higher transportation costs against cheaper (and often illegal) housing, the opposite to what one would expect from traditional residential location theory, framed from a first-world perspective (Alonso, 1964). According to the National Household Travel Survey conducted by Statistics South Africa (Stats SA, 2013) almost all households (98.9%) from the lowest income quintile in South Africa spend more than 20% of wages per capita on public transport (Stats SA, 2015). For the very poor of the developing world, whatever savings accrue from illegally squatting and living in squalor, are often negated by the expenses incurred to reach job opportunities as well as essential medical, educational and retail destinations (Cervero, 2013). The implication of this for modelling is that for these communities, the monetary cost of transportation may be more important than the generalized cost (see Section 4.2) and that the model must be capable of dealing with informal dwellings that are not part of the property market in the conventional sense.

Another challenge to the Government of South Africa, is striking a balance between investment in "economic infrastructure," such as transportation and energy supply systems; and "social infrastructure and services," such as housing, water, sanitation, health and a variety of grants/subsidies aimed at alleviating the plight of the poor without expecting a monetary return on the investment. Public transportation has attracted massive investments since the promulgation of the National Land Transportation Act of 2009. Like in most cities around the world, the new public transportation services (mostly Bus Rapid Transit) are subsidized, but probably more so as a service to the poor. During the first 20 years of democracy (1994-2014), the government has supplied approximately four million "housing opportunities" – 903,543 serviced stands and 2,835,275 houses or social housing units (Department of Human Settlements, 2019). Despite talks about more compact and sustainable cities, it seems that little has changed, because in the pursuit of helping as many households as possible, housing projects had to be built on affordable land, which could only be found on the outskirts. In a way, government has also made a trade-off between cheaper land and higher operating costs in the form of subsidies. This could still be mitigated by managing land use in such a way that densities are increased to achieve the necessary ridership levels to contain subsidies. The implication for modelling is that the government acts as a public-sector developer with important decisions to make about the location and size of proposed housing projects, with larger projects offering economies of scale but requiring tracts of land which will most likely only be found on the fringes. We revisit this as part of the Ekurhuleni case study (Section 7.2).

Yet another challenge is to ensure that public-sector investment contributes to economic growth and the reduction of persistent poverty and inequality. In 2011, the National Treasury introduced the Built Environment Performance Plan (BEPP) as a new statutory planning instrument aimed at achieving these elusive outcomes (National Treasury, 2015). In summary, the strategy is aimed at promoting economic activity in the urban hub of marginalized areas on the periphery, changing urban form by way

of walkable precinct plans, revitalizing the main activity area(s) in the city, linking these to the hubs by public transportation and promoting higher density, mixed-use development along the public transport corridors.

Lastly, it is important to note that ranked by GDP per capita (expressed in USD), South Africa occupies position 31 of 126 developing countries (Developing Countries Population, 2019). With a GDP per capita of about $6 600 compared to $40 000 to $80 000 for USA/Europe, South Africa is relatively poor with roughly 65% of the population considered to have a low income (Waldeck & van Heerden, 2017). This is the primary cause for the existence of informal settlements and has several implications for modelling residential and employment location choices as well as transportation mode choices.

3   Related work

In this section we explore possible similarities to, and differences from, other applications of UrbanSim in regions outside of the USA, where it was originally developed and implemented in several cities. A literature search found most applications in Europe (Schirmer, Zöllig, Müller, Bodenmann, & Axhausen, 2011; Patterson, Kryvobokov, Marchal & Bierlaire, 2010; Kryvobokov, Mercier, Bonnafous & Bouf, 2015; Di Zio, Montanari & Staniscia, 2010; Picard, de Palma & Kiarash, 2015) and Asia (Felsenstein, Lichter & Ashbel, 2014; Jin & Lee, 2018; Joo, Mehedy Hassan & Jun, 2011; Shi, Tong, Zhang & Tao, 2013). According to the classification provided by Developing Countries Population (2019), all these applications, except the one from China, were done in the developed world. The comparisons made to these application projects are done under the headings of *Country context* and *Data preparation*, which are for UrbanSim-SA further unpacked in Section 5.

3.1   Country context

Based on a substantive body of work following an informal European UrbanSim Users group meeting at ETH, Zurich in 2008, Felsenstein, Axhausen, and Waddell (2010) identified the following five differences in the prevailing land-use transportation environment between the USA and Europe (listed verbatim with comments regarding the situation in South Africa appended in italics).

First, a very different land-use environment exists in Europe. This makes for greater government regulatory controls over urban development at all levels. *Because of its political history, SA sees itself as a "Developmental State" with the need for statutory planning requirements such as the Built Environment Performance Plans, briefly described in Section 2. The government also acts as a property developer in the provision of free/subsidized housing to the poor (about 60% of all households (Rust, 2012)).*

Second, a very different attitude exists in European cities towards car ownership and dependence on public transport. *In this respect, SA is probably closer to the USA with 92% of white households owning one or more cars (Stats SA, 2018). Amongst black households, the figure is much lower at about 20%, mostly due to affordability rather than lack of aspiration. Public transport has received a lot of attention in the last decade but apart from Gautrain (fast intercity rail) and some new Bus Rapid Transport routes, public transport is viewed as something for those without other options.*

Third, Europe has a very distinctive ethos with respect to housing tenure and property rights. This results in levels of homeownership, composition of housing stocks and "acceptable" patterns of residential density, very different from those prevailing in the United States. *In South Africa most households aspire to own their homes, but ownership statistics vary between roughly 30% at age 30, to 70% at age 65 (Africa.com, 2018). Development densities have increased over time, but sentiment is very much against*

multi-story housing. While some of the policy alternatives presented in Section 7 have set targets for transport corridor densities as high as 400 housing units per hectare (hu/ha), corresponding actual densities at present are closer to 25hu/ha. With average (city wide) densities as low as 10 hu/ha, South African cities are even less dense than cities in the USA. The composition of housing stock also differs because about 18% of housing units in the metropolitan areas are informal (Stats SA, 2019b). Informal dwellings (sometimes referred to as "shacks") are mostly temporary, built from rudimentary materials and never compliant with building regulations. These can occur as one or more "backyard shacks" on a formal property (by arrangement with the owner) or as densely populated settlements, mostly on land that the occupants have no legal claim to (referred to as informal settlements).

Fourth, the major United States urban land-use issue, that of residential sprawl, features much less prominently in Europe where commercial (or employment driven) sprawl is high on the urban agenda. *In this respect, South Africa is more like the USA with residential sprawl high on the agenda, especially from the point of view of attempts to increase corridor densities to achieve the required ridership levels to contain public transport subsidies.*

3.2   Data preparation

Given that UrbanSim is a microscopic model that works with the attributes of every person, household, job and building in a city, it is well known that the process of preparing the input data for the base year of a simulation can be very time consuming. One publication from the list of applications in Europe and Asia (Schirmer, Zöllig, Müller, Bodenmann, & Axhausen, 2011) describes this process in great detail, pointing to difficulties associated with the collection of data from different sources and covering different periods of time, linking attributes to a common spatial unit for simulation, cleaning (for example by deleting duplicates) and imputing missing values. Surprisingly, this project (Zürich case study of SustainCity) was the only one based on the parcel geometry of UrbanSim, with all the rest based on grid cell and zone geometries.

One of the disadvantages of zone/grid cell geometries is that they are oversimplified representations of the real world, mostly containing multiple cadastral parcels and buildings, the attributes of which must then be aggregated or somehow accounted for. A common approach to achieve this is to create up to four synthetic buildings (residential, government, industrial and commercial) per zone/grid cell, for example (Felsentein et al., 2014) and (Patterson et al., 2010). The households obtained from a population synthesizer or small area statistics are then allocated to a zone/grid cell and from there to a single residential building. Even though this does not in itself aggregate or average household attributes like age of head, income or persons, the mere existence of households of any income group in the same building type is a form of behavioral aggregation. This could not possibly work in South Africa with one of the highest GINI coefficients in the world. For UrbanSim to succeed in South Africa, we adopted the parcel geometry from the outset and defined 5 income categories for sub-model segmentation of the Household Location Choice Model (see 4.1.3). We eventually used 10 residential building types (derived from settlement typology shown in Figure 2). The behavioral aggregation could, of course, be mitigated by creating more synthetic buildings (of different types) but it is not clear why one would introduce this additional complexity when parcel data is one of the most readily available datasets of all.

The literature review of related work has shown that while there are many similarities between the data preparation for applications in Europe/Asia and South Africa, there are two significant differences. Firstly, in South Africa we have found it necessary to avoid behavioral aggregation by using the parcel geometry, whereas all, except the Zürich case study in Europe/Asia, used zone or grid cell geometries. Secondly, all the applications in Europe/Asia, except for three (Kryvobokov et al., 2015), (Patterson et

al., 2010) and (Shi et al., 2013), had ready access to buildings data from some agency of government. In South Africa this was not the case and we had to acquire data from a private-sector company who derived it from satellite imagery (Geotera Image, 2013). Even if buildings data were available from an official source, it would not include informal buildings, which per definition bypass all municipal development approval processes. These buildings (approximately 15% of all buildings) would still have to be obtained from private-sector sources.

4   UrbanSim-SA

The development of an Urban Simulation Model for SA started with a review of the state of the art in 2007. At the time, the objective was to develop a model that would be capable of representing cities as complex, adaptive and self-organizing systems, as described by Batty (2005). The study concluded that approaches based on understanding the behavior of agents, which closely resemble actors in everyday life and relying on the behavior of the overall system to emerge from the interactions between agents, had the best chance of achieving the objective. UrbanSim was thus selected as the core of the Urban Simulation Model.

A proof of concept phase then followed between 2009 and 2012 in which the Department of Science and Technology funded the development and application of the model in the cities of eThekwini, Nelson Mandela Bay and Johannesburg. This experience provided an affirmative answer to the biggest question at the time: whether it would be possible to collect the attributes of the 5 main tables required by UrbanSim (parcels, buildings, households, persons and jobs) in a sparse data environment, with no building data available from any of the cities. This, together with some behavioral validation, was enough to secure further funding to address outstanding issues such as the place of work (see Section 5.6) and to complete UrbanSim-SA.

For readers interested in the progression of ILUT models since 1960, a paper by Moeckel (2017) provides an excellent overview. Although his literature review was done from the perspective of determining how ILUT models deal with constraints that may influence household location choices (such as price of dwelling, travel time to work or monetary commute cost) it provides a comprehensive list of references, including those considered by the CSIR in 2007. Just out of interest, it seems that apart from SILO (Ziemke, Nagel, & Moeckel, 2016), no completely new models were developed since 2007—when the CSIR selected UrbanSim.

4.1   UrbanSim

A very brief description of UrbanSim is provided in this section to define the context against which it was adapted as described in the next section. For more comprehensive descriptions, please refer to Waddell (2002, 2011).

UrbanSim is a microscopic land-use model in which the attributes of every person, job, household and building in the city are represented by a separate record in the database. The persons, jobs, and households for the base year of simulation (which must be a census year) are generated by a population synthesizer, as described in Section 5.1. During the simulation, the base year population and jobs are adjusted for each subsequent simulation year by the Household and Employment Transition Models, respectively. These models grow the population and jobs each year according to city-wide control totals. Other options are available for simulating various life-stage events. The spatial resolution can also be microscopic, at individual property boundaries, but zone-based geographies can also be used.

UrbanSim recognizes the property market as an important construct of the urban system and ac-

counts for the current stock of real estate (and its spatial distribution) as well as the supply and demand of new stock. Households and businesses represent the consumers of residential and non-residential stock, respectively. The choice of location by these agents is accounted for by the occupation of space in a building.

Developer agents, representing the supply side, use occupancy rates as a feedback signal to determine the rate at which new stock can be built and sold. Supply and demand are mediated by price, another feedback signal that influences the return on investment calculation done by developers when considering/choosing between potential developments.

Governments set policies that either regulate the use of land or influence development through pricing policies such as developer contribution fees. Governments also build infrastructure and provide social facilities such as libraries, clinics, community halls, parks, sports fields, and fire stations. Transportation infrastructure and the quality of social facilities influence the attractiveness of locations for different consumers. In UrbanSim-SA, we treat government housing projects as development events, each representing a project built on a specific parcel of land. The location and number of housing units per project are the product of a comprehensive planning process (not market forces) but the total number of units delivered per year is invariably limited by budget constraints.

Households have characteristics that influence their preferences for housing of different types at different locations. Similarly, businesses have preferences that vary by industry/economic sector and number of employees for alternative building types and locations. Two models are responsible for modelling these choices, the Household Location Choice Model and the Employment Location Choice Model. Both are discrete choice models based on the pioneering and Nobel Prize (2000) winning work of McFadden (1974) on Random Utility Maximization Theory. This approach determines the probability of a choice from a set of available alternatives based on the characteristics of the chooser and the attributes of the alternatives, depending on the relative utility that an alternative offers the chooser.

The systematic component of a utility is expressed as a linear combination of estimable coefficients multiplied by independent alternative-specific variables that may be interacted with the characteristics of the agent making the choice. The coefficients are estimated using maximum likelihood methods built into UrbanSim. For choice models, the estimation can be based on the stated or observed behavior of agents. All our work so far has been based on observed behavior derived from sources such as the census and household travel surveys.

Examples of alternative-specific variables often used in utility calculations include the distance, travel time and generalized cost of commuting. The term generalized cost refers to the sum of the monetary cost and the value of time spent undertaking the trip, which varies according to the traveler's income and the purpose of the trip.

The process of selecting explanatory variables usually starts by including some variables that are consistent with urban economic theory. Guided by various indicators which estimate the relative importance and statistical significance of each variable, variables are added or deleted to maximize the overall explanatory power of the model. Some judgement may be required to avoid situations like the following: If distance to employment is used as a variable in the Household Location Choice Model, then estimating the sub-models for low-income groups in a South African city will probably show that the variable is significant and that these households prefer to live far from work. Since this is more likely a remnant of apartheid than an actual observed preference, the simulated future city might be no different from the present. In such cases it may be necessary to omit the variable, define a more appropriate variable or estimate the model on a subset of data that will not perpetuate the situation.

4.1.1   Adapting UrbanSim

For this paper, adaptation is defined as additions or modifications to/of the source code or configuration files that were necessary for UrbanSim to succeed in South Africa.

4.1.2   Data required for base year

By far the bulk of the effort was consumed by the development of 6 completely new models needed to generate microscopic building data and to place households and jobs into specific buildings for the base year of the simulation. The most important of these are discussed in Section 5.

4.1.3   Affordability constraints

In a review of ILUT models, Moeckel (2017) found that most models do not explicitly account for affordability constraints in the location choices made by households. While this also applies to UrbanSim, users around the world have partially solved the problem by introducing an interaction variable based on the product of household income and the unit price of dwellings. This raises the probability of lower income households choosing cheaper dwellings (and vice versa) but is not quite the same as a constraint.

From the outset, the Household Location Choice Model was set up to have 5 sub-models, one for each household income category. The advantage of such a configuration is that it allows completely different explanatory variables to be used for each income category, something that was expected to be important in South Africa with one of the highest GINI Coefficients in the world (World Bank, 2011).

Despite these configuration changes, early results indicated that there were far too many inappropriate choices of rich households living in shacks and poor households living in mansions. The problem was then solved by coding a filter, which ensured that the 30 alternatives, from which a household would make a choice, were all sampled from buildings that were affordable to households belonging to a specific income category.

4.1.4   Proximity to employment

Most UrbanSim models around the world are configured to use various measures of proximity to employment as explanatory variables in their Household Location Choice Models. The measures could be based on travel distance or time/cost of travel by various modes of transport. The explanatory variables are usually expressed as a weighted average or logsum of these measures from the location of a dwelling being considered by the household to all jobs. By this definition, any dwelling close to an area with a high job density, will present with a relatively high utility. If the household chooses one of these dwellings, which could be far from the current place of work, the plan is probably to then look for a job closer to home.

In South Africa, where employment is scarce (with official unemployment close to 30% (Stats SA, 2019c)) fewer households would relocate before first securing a new job. A more realistic representation of this behavior in the model would therefore require the measure of proximity to be based on the travel time/cost from the dwelling being considered to the existing job rather than the travel time/cost from

the dwelling being considered to all jobs.

The implementation of this change was complicated by the fact that there could be more than one employed person per household. Although not ideal, it was decided to take a first step by calculating a travel_cost_to_work variable (by mode of transport) for the head of the household.

4.1.5   Demographic transition

Once the travel_cost_to_work variable had been implemented, a problem with the demographic transition models became apparent. The Household Transition Model was found to maintain the relationship between new households and new persons created every year, but the Employment Transition Model created new jobs with the correct employment sector but with a person_id of zero (meaning that the job was not yet linked to a person).

Since none of the new jobs (which eventually become most jobs) could be linked to a household via the person_id, it was not possible to calculate the travel_cost_to_work variable for many households. The solution was to code a new model, which randomly linked the new jobs to new persons with an employment status of *employed* or *unemployed but seeking employment*. All unassigned jobs were then deleted because the Employment Transition Model would maintain the correct number of jobs per employment sector in the next simulation year. None of this affected the work-location of the new jobs.

4.1.6   Excess capacity

A last feature that required modification is that when development project proposals (for construction of new buildings) are evaluated, large projects tend to yield the highest return on investment and often get selected, leading to land parcels being completely built up in one year. This is no problem in proclaimed townships where farms/smallholdings have already been subdivided. Several cities, however, contained massive farm parcels on the periphery. If these parcels were built up, it had a substantial impact on future development because the high vacancy rate for that type of development inhibited similar developments from taking place elsewhere in the city (possibly for many years). The fact that the parcel was fully built up also prevented buildings of a different type from being constructed on the parcel in future years.

An elegant solution to this problem would have been to subdivide the large parcels into something that resembles proclaimed townships, based for example on procedural city modelling (Lechner, Watson, Tisue, Wilensky, & Felsen, 2004) or example-based texture synthesis (Vanegas et al, 2009). While this would be desirable for 3D visualization of growth, another solution was adopted for computational benefits, which simply involved demolishing the excess capacity at the end of every simulation year and releasing the associated land for future development.

4.2   Transport model

The transport model used with UrbanSim-SA is described in detail in Waldeck & van Heerden (2017). For this paper we only provide a short summary of how an adapted version of OpenTripPlanner (OTP) can be used to calculate the lowest monetary cost of commuting between pairs of origin-destination zones for any of the available modes of transport during the morning peak. Readers interested in the rationale behind using monetary cost and distance variables as proxies for generalized cost variables (which depend on the value of time and congested state of the network) are referred to section 4.3 of Waldeck & van Heerden (2017).

Various data sources are required to run OTP, including a road network, origin-destination zones, public transport data, and the cost of private vehicle usage. A short description of each follows.

4.2.1   Road network

The road network was generated from OpenStreetMap (OSM), using the Geofabrik GmbH (2018) servers to obtain the file. In the case of simulating scenarios that involved future changes to the road network, typically driven by the Provincial Department of Transport, the OSM network was edited to reflect the planned changes; new roads were drawn in using JOSM. A new network file was provided as an input to OTP and the output, in the form of lowest cost commute options by mode, provided to UrbanSim for use in simulation years following the network change.

4.2.2   Origin-destination zones

To simplify matters, the origin-destination zones were identical to the analysis zones introduced later in Section 6. The lowest-cost trip for any pair of origin-destination zones (with area of about 3km²) were based on the network distance between the centroids.

4.2.3   Public transport network

The public transport network was created from public transport itinerary data, which were obtained from the respective municipal transport departments. The raw data, in various formats ranging from Excel sheets to PDF files, were converted into the General Transit Feed Specification (GTFS) format and subsequently used in OTP. OTP combines the public transport network with the road network and builds a directed graph to be used in subsequent routings.

Due to data limitations and the complexity of the minibus taxi industry, minibus taxis were included as a means of public transport with stops distributed at 2km intervals along their routes, since these vehicles tend to stop anywhere along a route to pick up or drop off passengers. Taxis are licensed to operate on one or more routes, which were available from all 3 cities. The routes were not changed during the simulation period because route-planning is known to occur *ad-hoc*. The schedules, which include departure times and frequencies, were also obtained from the cities.

4.2.4   Private vehicle usage

Private vehicle usage (also referred to as the *drive alone option*) was included by means of a fixed, per-kilometer cost, based on rates published by the Automobile Association. The calculations are described in Waldeck & van Heerden (2017).

4.2.4.1   Determining lowest-cost commute trips

Given the input data, the final step was to determine the lowest-cost commute trips between pairs of zones. This was achieved by adapting the Batch Processor in OTP to calculate the cost between origin-

destination pairs rather than using shortest travel time, as is the case in most developed countries. A walking radius of 2km was used for people to access public transport and walking and cycling were modelled as not incurring a monetary cost.

4.2.4.2   Update frequency

In most cases OTP was run in the base year (2011), and then re-run for a later year only if there were substantial changes to the road or public transport networks that needed to be taken into account in future simulation years. In such cases, OTP was run again and the output provided to UrbanSim in the form of the usual travel_data table but in a separate folder from the base year so that parameterized variables would read from the correct folder for the year being simulated. Note that even in the simplest case of using the same travel_data for all years, the proximity to employment measures discussed in Section 4.1.4 are different for each year because the variables are updated according to the constantly changing location of households and jobs.

5   Data preparation

One of the challenges that come with using the parcel version of UrbanSim is that it requires microscopic data about every person, household, job, building and land parcel in the city for the base year of a simulation.

Much of the effort in preparing the base year data goes into establishing the relationships between these entities. For example, households, persons and jobs are obtained from the most recent census. A household contains one or more persons, some of which may be employed (have a job). These relationships are easy to establish because households, persons and jobs share the serial number of the same enumeration form. But which households live in which (mostly residential) buildings and which jobs take place in which (mostly office, retail or industrial) buildings?

The following sections describe the most important attributes of these entities and how the relationships were established in a sparse data environment characteristic of developing countries.

For readers that may be embarking on a journey with UrbanSim, several publications point to the fact that the data preparation phase is iterative. It is not uncommon to discover incorrectly classified parcels during the validation phase, taking one back almost to the start of the workflow. For this reason, it pays to automate as much of the process as possible, see for example (Schirmer et al., 2011). In our case we achieved similar objectives by extensive use of SQL and ArcGIS ModelBuilder scripts.

5.1   Synthetic population

The household and job agents used by UrbanSim were derived from a 10% sample of enumeration forms from the last census (2011) by a technique known as Iterative Proportional Updating (Ye, Konduri, Pendyala, Sana, & Waddell, 2009). This explains why the base year must always be a census year. The Population Synthesizer supplied with UrbanSim, required minor modification because the enumeration form used by Stats SA differs substantially (in format and content) from the form used by the US Census Bureau. The modifications were limited to renaming data fields and the order/format in which they were read in.

In terms of spatial resolution, Stats SA's sub place geography was found to provide the closest match to the average population of the BLOCK GROUP geography of the US Census Bureau, which was in turn used for the development and testing of the Iterative Updating Algorithm.

During the simulation, the population and jobs are adjusted by the Household and Employment Transition Models, which grow the households and jobs each year according to a set of exogenous control totals. The city-wide control totals for every simulation year were provided by IHS Markit (formerly known as IHS Global Insight) from their Component Cohort Model. Household and population totals were classified by race, income, and age groups. Employment totals were provided by one-digit Standard Industry Classification.

5.2   Parcel attributes

To use UrbanSim at a resolution of property boundaries, one needs to categorize the following attributes for each land parcel: land-use type, development template, improvement value and land value. Several other attributes are required to restrict development where parcels are in wetlands, on steep slopes, etc. These are simple to apply except if restrictions such as flood lines cut through a parcel so that half of the parcel can be developed and the other half not. To avoid this, the data preparation workflow, at an early stage involves an overlay of the parcel layer with restriction layers such as flood lines, hydrology and geology. This allows any part of a parcel within a flood line or pond/wetland to be labelled as "non-developable" by a script in the GIS.

Municipal valuation rolls in South Africa provide a market-related value for parcels including all buildings on the parcel, without distinguishing between the value of the land and the value of the buildings (the so-called improvement value). This was resolved by introducing the notion of a "development cost," which includes the improvement value and the value of the land associated with each building on the parcel, as an attribute of buildings rather than parcels.

A critical step in enumerating the remaining attributes was the selection of a settlement typology based on a cluster analysis done by the Knowledge Factory (2006) on factors including socio-economic rank (income, property value, education and population group), life stage (age, household and family structure) and dwelling type (size, type and age of structure). The analysis, originally aimed at geo-demographic segmentation for micro-marketing purposes, identified 10 clusters comprising 38 classes, represented in Fig. 1 on axes of income and development density.
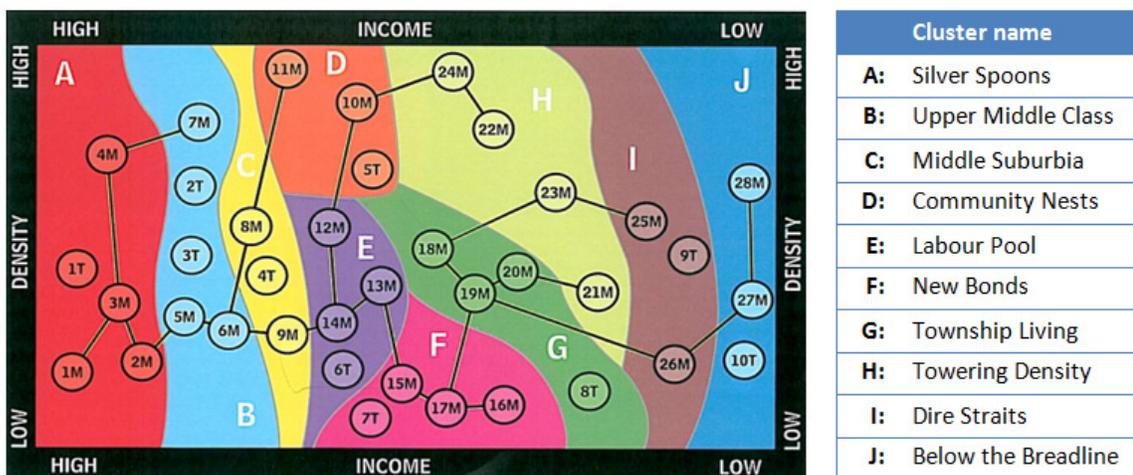
Figure 1. Settlement typology derived from geo-demographic clustering of (mostly) census data by the Knowledge Factory, originally for micro-marketing purposes

A few examples of the 38 classes in the typology are shown in Fig. 2, including four of the five classes in the Silver Spoons cluster, two of the three classes in the Dire Straits cluster and two of the three classes in the Below the Breadline cluster.

1M: Upper Crust

They are the elite of South African society – anointed through wealth and achievement. The Upper Crust live lives of distinction in pampered luxury, with little concern about cost – for them, quality is not negotiable. They know and can afford the best the world has to offer. Properties

2M: Pearl Strings

Closely related to the Upper Crust, the Pearl Strings are crowning lifetimes of achievement with refined, slightly understated style – in fact, they may well frown on flash. While their incomes are only outstripped by those of the Upper Crust, properties

3M: Cheese and Wine

They insist on the American Dream in South African suburbia, on being the captains of their own ships and on recognition of their status as self-made. A decade or two ago, the Cheese and Wine led the trek from the older suburbs to the new suburban Meccas

4M: Fashion Café Society

They are hip and happening – the trendsetters that push themselves hardest to live according to the dictates of lifestyle magazines. They work hard, earn big and, sometimes, spend even bigger. Fashion Café Society represents the new wave of residents of

25M: Chakalaka

Chakalaka clusters (named after a spicy vegetable relish/dish developed in the townships of Gauteng) were meant to be orderly locations – much like the eKasi clusters, however, all open spaces in this cluster have been crammed full with a wide assortment of shacks and structures. The result is a lively community that is,

26M: Poor Neighbours

The residents of the Poor Neighbours cluster, too, have outgrown the old 'matchbox' houses originally built in the area. As a result, the cluster is typified by numerous shack dwellings erected amongst the permanent structures or nearby. Dwellings are basically standard four-room or three-room

27M: Tin Town

When the people of the Tin Town cluster go to bed at night, a good dream would be to wake up somewhere else for it is difficult to find redeeming factors about this most oppressive cluster, other than the tenacity of its inhabitants. Tin Town clusters consist of very dense, relatively small shack settlements. What distinguishes this cluster

28M: eKaya

Informal settlements are nothing new in South Africa – eKaya clusters are proof of this. They are, however, different from newer informal settlements, being older, with the majority older than a decade and, often quite far from the city centres. Properties are slightly larger and more established – you might well find a clearly

Figure 2. Examples of the settlement typology (8 of 38 classes) derived from Knowledge Factory Cluster Plus product with abbreviated description and representative photo of each class.

From the outset, the 38 classes of the settlement typology were used as development templates. The constructs of development templates and corresponding development template components allow UrbanSim to configure virtually any development proposal for construction in the future. The template structure is robust enough to cater for anything from a single house on an infill lot to a mixed-use project with retail on the ground floor and apartments above. Backyard dwellings were for example represented by one formal building and one or more informal buildings on a parcel with a formal development template. These were limited to land-use types on which backyard dwellings have been observed in the past.

While the development templates worked well, it proved exceptionally difficult to find suitable predictors of the market value of a property until the clusters were introduced as land-use typology. This was inspired by the observation that the cluster analysis done by the Knowledge Factory was based on factors which included household income, property value and dwelling type, so for residential properties at least one would expect to find a correlation between the newly defined land-use type and property value, which turned out to be true.

A potential disadvantage of using the clusters as a land-use typology was that it deviated from what municipal officials are used to. Fortunately, UrbanSim allows for the definition of several typologies in addition to development templates and land-use types. The plan type is one such typology that is used to define land use in the way that planners are accustomed to but in interactions with metros it seemed that most were indifferent to the definition of land use while one even welcomed it as a new way of thinking about their own data and processes.

5.3   Building attributes

The number of jobs that can be accommodated in a building depends on the floor space available in the building and the floor area required per job per sector of the economy. The total floor area is related to the total parcel area through the so-called "floor area ratio." The number of households that can be ac-

commodated on a parcel, which determines the density typically expressed as housing units per hectare, also depends on the floor area per residential unit and the floor area ratio. The market value per residential unit may also depend on the floor area and parcel area.

Unlike in the developed world where much of this information is available from municipal building records, we have to obtain the type of building from one of 70 classes of the "GTI Building based land-use type" dataset (Geoterra Image, 2001, 2011) and the rest from observed average densities per building/land use/development template type. This involves a fair amount of analysis, for example, to exclude outliers caused by building projects in progress from the calculations.

5.4   Business entities

Information about business entities in South Africa is severely lacking and the best that can be done is to use jobs as a proxy for the behavior of businesses. Pseudo-buildings are created with enough floor space to accommodate the estimated number of jobs per economic sector but there is no way for example to distinguish between businesses trading in clothing from businesses trading in fast foods, which prevents the modelling of agglomeration behaviors.

5.5   Relationships

The most challenging relationships to establish are those between households, jobs, and buildings. In the case of households, the synthesizer specifies the sub place where every household resides. The households within each sub place, about 1600 on average for the three major metropolitan municipalities in Gauteng according to the 2011 Census (Stats SA, 2019a), are allocated to specific buildings in that sub place on affordability considerations derived from the income of the household and the municipal valuation of the pseudo property.

In the case of jobs, the relationship between jobs and households is easily obtained by linking jobs to all the employed persons in a household. Since households are located, from census information, this linkage provides the place of residence of the person holding the job but the census (since 2011) provides no information whatsoever on the place of work of the person.

5.6   Place of work

The place of work issue was resolved by a deceptively simple method based on using building valuations as a proxy for the number of jobs that a building can accommodate combined with allocating jobs to buildings in such a way that the home-to-work travel distance distribution for the city as a whole matches the last municipal travel survey. Valuation rolls are generally of high quality because they are required by law as the basis for taxation and open to dispute by any resident.

The job capacity of a building is assumed to be proportional to its valuation by a factor used to compensate for the number of jobs per million of the valuation for different types of buildings. This factor serves as compensation for differences in the contribution of land and buildings to the overall valuation, different methods of construction and different job densities resulting from different activities in the building. Considering the method of construction, a typical industrial building with steel frame is much faster and cheaper to build than a multi-story concrete building with luxury finishes characteristic of shopping malls. As an example of how business activities affect job density, consider a warehouse that should have a comparatively low job density because most of the space in the building is used for the storage of goods.

The building-type correction factors were determined empirically through 3 to 4 iterations of the following process:

1. Multiply a copy of the municipal building valuations by the building-type correction factors for the following types of buildings: retail, commercial, industrial, institutional (mostly government, correctional services, defense, emergency services and police) and other (utilities, community services, health care, education, recreation). The correction factors are set to unity for the first iteration.

2. Disaggregate the total number of jobs in the city (from exogenous employment forecasts provided by IHS Markit, 2012, 2016) to individual buildings in proportion to the building valuation divided by the total valuation of all non-residential buildings in the city.

3. Compare the total number of jobs by building type to the employment forecasts by job sector, adjust the correction factors and iterate.

The last step requires the highest-level building-type classification to match the employment sectors, achieved through combining sectors such as "agriculture" and "electricity, gas and water supply" into "other." In a few cases it was necessary to introduce correction factors at sub-classifications of building type. Regional malls for example were found to have unrealistically high job densities probably due to the cost of prime land, method of construction (multi-story concrete) and luxury finishes. Another example is sports stadia, which are extremely expensive to build, yet used infrequently with very low job density. Sports facilities at universities and schools were also found to contribute significantly to the valuation but detract from job density because they employ relatively few people for maintenance purposes.

Once the building-type correction factors have been determined, the relationship between the place of work and the place of residence of the person holding a job is determined by first placing home-based jobs (about 12% for the City of Tshwane, according to 2011 Census (Stats SA, 2019a)) in the same building as the place of residence. The remaining jobs are then placed so that the overall commute-distance distribution (including home-based jobs) has a median as close as possible to the median distance of work-related travel from the last household travel survey commissioned by the municipality (about 14km in the City of Tshwane).

6   Model validation

Empirical validation was done by simulating a period in the past (2001 to 2011) and comparing the growth forecasts with the actual growth that occurred during the same period according to two reputable sources, Stats SA and Geoterra Image (GTI).

The analysis zones used for the comparison were derived from the 2001 sub places published by Stats SA by retaining or merging smaller sub places in built up areas and subdividing larger sub places into areas of about 3km². This was necessary to avoid complications that arose during the proof of concept phase from the Modifiable Areal Unit Problem when growth comparisons were based on areas that differ significantly in size. The solution required the household counts (2001 and 2011) to be adjusted by dasymetric mapping using the dwelling counts of GTI as a proxy for how the population is distributed within each sub place. A similar correction would have been required anyhow because Stats SA changed the sub place boundaries between the 2001 and 2011 censuses. The resulting analysis zones (varying in number between about 800 and 1000 for the three case studies) are shown in Figure 3.
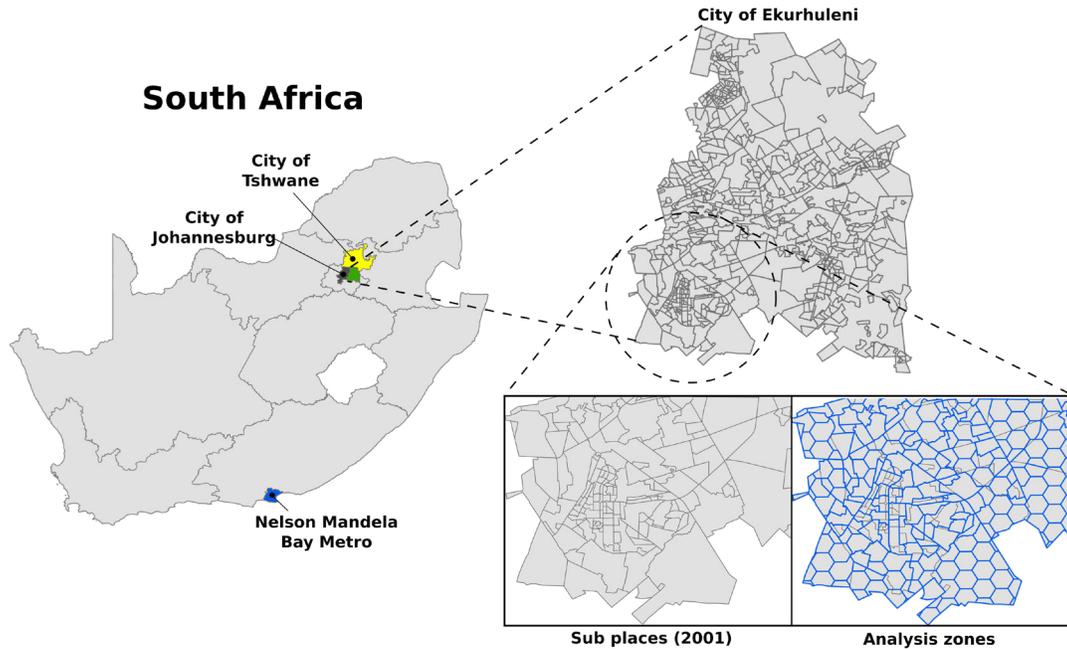
Figure 3. Analysis zones with near-uniform area of approximately 3km^2 for the city of Ekurhuleni, illustrating how small (densely populated) sub places were merged and large (sparsely populated) sub places were split to produce the analysis zones

The results of the empirical validation are shown in Figure 4 as a comparison of the growth in households forecast by the Urban Simulation Model to the actual growth in households according to GTI between the 2001 and 2011 censuses as well as a comparison of the actual growths according to Stats SA and GTI (right). Upon investigation of the top outliers produced by UrbanSim, very few were found to not have logical explanations. Even with the outliers, Figure 4 suggests that the predictive accuracy of the model is comparable to a comparison of the actual growth measured by different methods employed by two reputable organizations. While it is conceded that there is a 1-year difference between the observations of Stats SA and GTI (2013 release based on 2010 remote sensing), we regard the results as a satisfactory overall validation not only of the Urban Simulation Model but of the preceding data preparation processes.
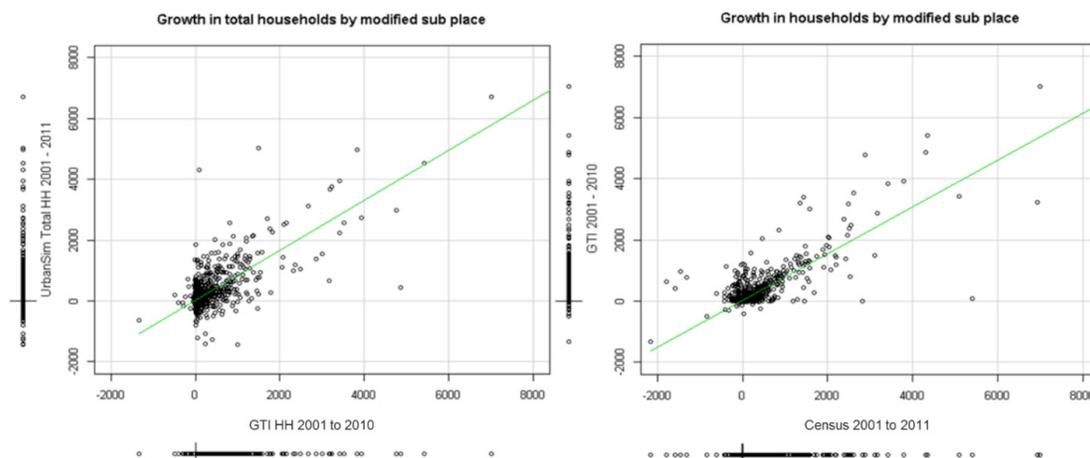
Figure 4. Model validation by comparison of UrbanSim results with actual growth in the number of households per analysis zone according to GTI (left) and actual growth according to GTI with actual growth according to Stats SA (right) over a similar period in the City of Tshwane

7 Results

A sample of results obtained from simulations done for the Cities of Tshwane, Ekurhuleni, and Nelson Mandela Bay between 2013 and 2017 is presented next. In each case a brief description is provided as context, followed by a list of the policy scenarios that were simulated (in addition to the Trend Scenario defined as what most stakeholders in the participation process would regard as a given), followed by sample results for one of the scenarios and a short discussion.

7.1 City of Tshwane

7.1.1 Context

For the City of Tshwane, the objective of simulating various policy scenarios was to inform the development of their Capital Investment Framework (CIF) with specific reference to quantifying the future demand for services rendered by government through social facilities as a means of determining the capital budget requirements of such facilities.

7.1.2 Scenarios

In this instance, potential disagreements about what constitutes each of the scenarios were avoided by not attempting to specify all the scenarios at the outset, but to first focus on the Trend Scenario. This scenario involves the fewest assumptions because it simply represents what most stakeholders in the participation regard as a given, for example that the principles outlined in the Regional Spatial Development Framework (RSDF) will be adhered to and that all phases of the *Integrated Rapid Public Transport Network* (IRPTN) will be implemented.

    The other scenarios were developed later from the interactions that followed the presentation of

results obtained from simulation. In this way it was not too difficult to specify and simulate the following scenarios:

1. Trend scenario with most likely city-wide demographic and employment growth projections provided by IHS Markit in 2012 (since used extensively by various departments)
2. Pessimistic socio-economic growth scenario with higher population and lower employment growth projections provided by IHS Markit in 2016 (taking a dimmer view of regional migration and of the South African economy)
3. Priority areas scenario advocated by engineering groups based on the elevated cost of providing wastewater infrastructure in some catchments.

7.1.3   Model estimation

In this section we provide the results of estimating 3 of the 5 sub models of the Household Location Choice Model (HLCM) for the Priority areas scenario to provide some insight into the explanatory variables that were found to be most significant.

Table 1. Estimation results for 3 HLCM sub-models for the priority areas scenario

Sub model	Explanatory variables	Estimate	Standard error	t-values
1 (low income)	b_ln_nhb_jobs_within_20_km	0.67972	0.00190	357.7
	b_ln_proxim_ind	0.08575	0.00081	105.6
	h_x_b_transit_travel_cost_to_work	-1.77552	0.01747	-101.7
3	b_logsum_access_da	0.80883	0.00067	70.5
	h_x_b_transit_travel_cost_to_work	-1.17697	0.00813	-162.6
5 (high income)	b_hwy_3000	0.57299	0.01842	31.1
	b_ln_avg_inc_zone	0.03791	0.00312	12.2
	b_ln_empden_zone_6	0.09605	0.00475	20.2
	h_x_b_da_travel_cost_to_work	-0.06702	0.01465	-4.6
	h_x_b_transit_travel_cost_to_work	-0.39081	0.02269	-17.2

In Table 1 above, all variable names starting with "b_" are building variables (returning a one-dimensional array with a value for each building) while the remaining two are interaction variables returning a two-dimensional household x building array. A brief definition of the variables follows:

- b_ln_nhb_jobs_within_20_km: Natural logarithm of the number of non-home-based jobs within a 20km network distance of the building.
- b_ln_proxim_ind: Measure of proximity based on the aggregated area of industrial land within the analysis zone that the building is located in.
- b_logsum_access_da: Logsum of travel cost by "drive-alone" mode of transport between the zone that the building is located in and all other zones.
- b_hwy_3000: The building is located within 3km from a highway. In this case not a network distance, obtained from simple GIS calculation.
- b_ln_avg_inc_zone: Natural logarithm of the average income of households in the analysis zone in which the building is located.
- b_ln_empden_zone_6: Natural logarithm of the density of jobs in the trade and financial services sectors of the economy within the analysis zone in which the building is located.
- h_x_b_da_travel_cost_to_work: Natural logarithm of the monetary cost of travel between the analysis zones representing the place of residence and place of work of the head of a household

for the "drive-alone" mode of transport. See section 4.1.4 for further discussion.
- h_x_b_transit_travel_cost_to_work: Same as above but for the "transit" mode of transport.
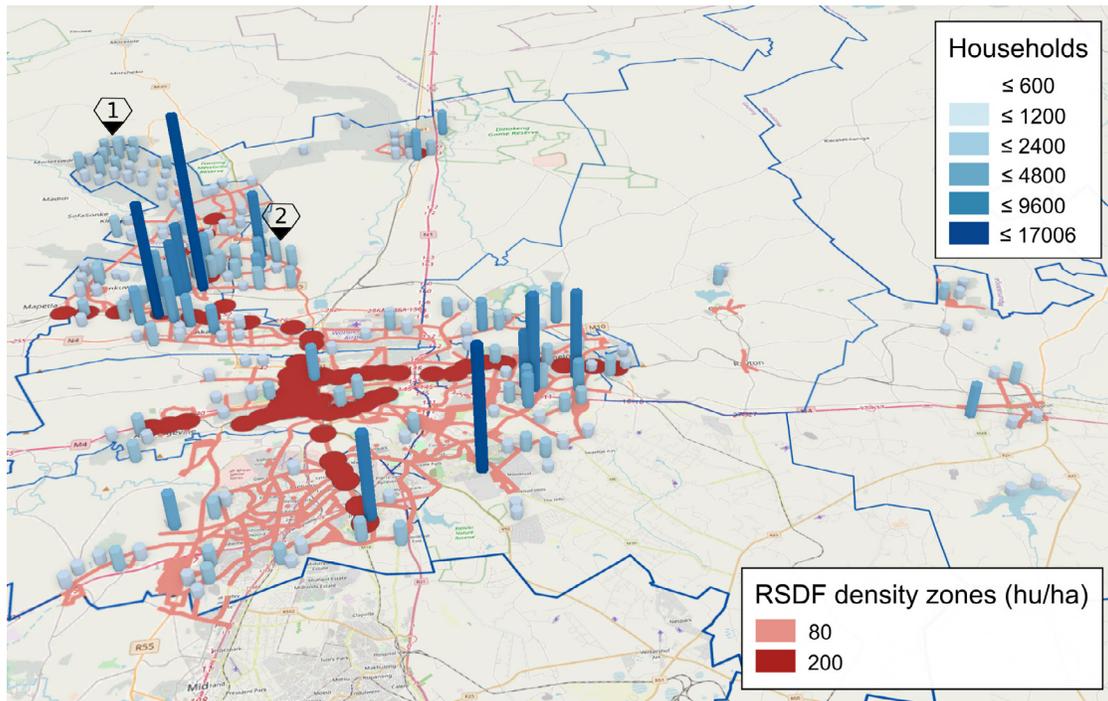
Figure 5. Trend scenario; the height of each bar represents the growth in number of households per analysis zone between 2011 and 2030
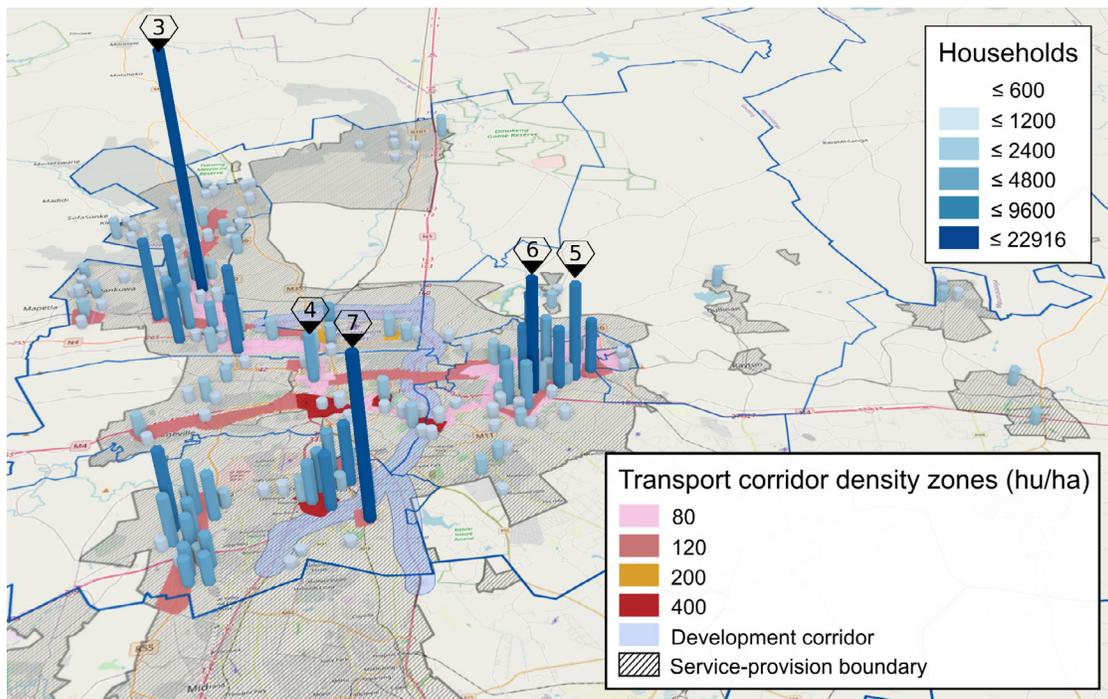
Figure 6. Priority areas scenario; the height of each bar represents the growth in number of households per analysis zone between 2011 and 2030

A comparison of Figures 5 and 6 (showing only the growth in households between 2011 and 2030)) illustrates just how important the densification zones, as an expression of strategic intent, are in shaping the spatial form of the city. The two scenarios used identical demographic and employment projections and no changes whatsoever were made to the model specification that could have influenced where households choose to live or work. Yet the Priority areas scenario (Figure 6) anticipates a more concentrated development pattern that is better aligned with public transport corridors than the linear densification zones promoted by the RSDF in the Trend scenario (Figure 5). Consider for example the absence of development in the area corresponding to marker 2 in Figure 5. The absence of growth in the Winterveldt area (marker 1 in Figure 5) can be ascribed to the priority areas including a services-provision boundary, beyond which it becomes too expensive for the city to provide bulk wastewater services. The growth at marker 7 can be attributed to the release of a large tract of land for green-fields development in proximity to the N1 freeway.

Figures 5 and 6 also illustrate the impact of government-funded housing projects with significant growth projected at these sites (markers 3, 4, 5 and 6). In this case study, the location of the housing projects resulted from a prior BEPP planning process (alluded to in Section2) with the number of housing units built per year limited by the available budget. In the next case study, we show how such location decisions have been investigated to inform future BEPP planning processes.

7.2   City of Ekurhuleni

7.2.1   Context

For the City of Ekurhuleni, the objective of simulating spatial growth patterns was to understand the impact of various interventions. These included catalytic projects, earmarked to stimulate economic growth; to determine how to best unlock the economic benefits from implementing the Aerotropolis Master plan, and to determine the trade-off between large government-funded housing projects and smaller in-fill development.

7.2.2   Scenarios

As was the case in the City of Tshwane, the first instance involved only the principles set out in the Municipal Spatial Development Framework (MSDF) and only projects that were certain to be implemented. These projects included both housing projects and new roads and an expansion of the public transport network.

The other scenarios were again developed from the interactions that followed the presentation of results obtained from the trend scenario simulation. The following scenarios were identified:
1.   Trend scenario with most likely city-wide demographic and employment growth projections provided by IHS Markit in 2016.
2.   Aerotropolis scenario with the same demographic projections as the Trend scenario, but higher employment projections as indicated in the *aggressive option* of the Aerotropolis Master Plan.
3.   Housing projects scenario that sought to quantify the impact of accommodating about 100 000 households qualifying for free government housing in smaller projects located as close as possible to employment opportunities.
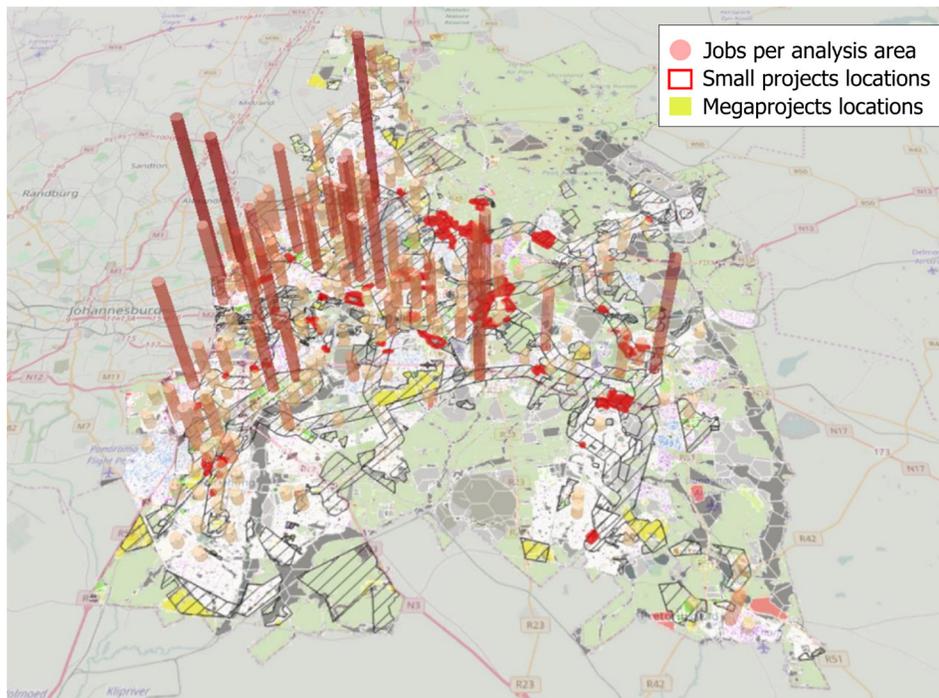
7.2.3   Results

Figure 7. Trend scenario; proposed location of large housing projects in relation to number of jobs per analysis area in 2011
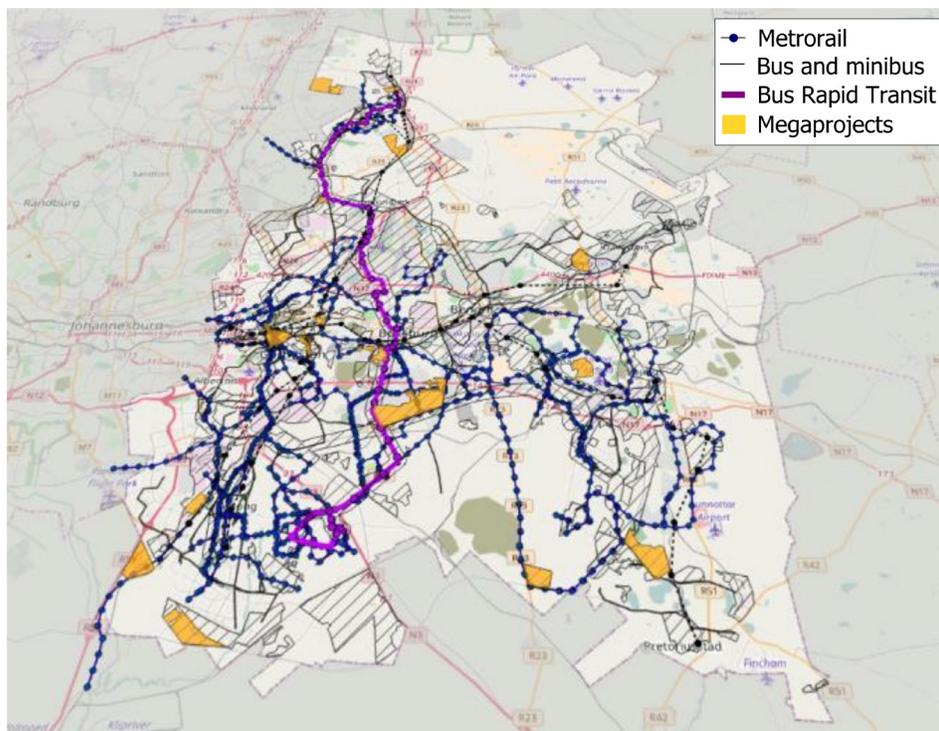

Figure 8. Public transport network in 2011

The location of the 139 best vacant parcels to accommodate the 100 000 households on the basis of their accessibility to employment are indicated in red in Figure 7 in relation to the number of jobs per analysis area in 2011 and the location of the proposed large projects (yellow in Figure 7 and orange in Figure 8).

As a first step towards facilitating debate between departments, *impact* was defined simply as the average monetary commute cost of the 100 000 affected households in 2030, based on the actual place of residence and place of work for each household. A more comprehensive definition of impact would have to consider many other factors such as the cost to government and the cost to the environment. The cost to government would have to at least consider the capital and operating cost of bulk infrastructure and transit.

Based on the simple definition the smaller projects option only reduced the average commute cost by an insignificant 5%. Upon investigating this counter-intuitive result it was found that the apparently remote larger projects also had good accessibly to employment due to the well-connected and subsidized public transport network illustrated in Figure 8.

7.3   Nelson Mandela Bay

7.3.1   Context

The objective was to analyze and compare long-term growth patterns that could result from the implementation of specific spatial planning scenarios and how the resulting demand for services would impact on the Long-Term Financial Sustainability Plan being developed by the city.

7.3.2   Scenarios

Developing the spatial planning scenarios was simplified considerably by the fact that the project started shortly after the city had gone through a scenario development process through which a "Walking Together for Growth" scenario was adopted as the preferred future for the city. Recognizing that financial sustainability involves much more than spatial planning and demand forecasting, the following scenarios were simulated to enhance the city's prospects of achieving the desired future state:

1. Trend scenario with most likely demographic and employment growth projections provided by IHS Markit in 2012.
2. Trend scenario with more optimistic demographic and employment growth projections which would be more characteristic of the "Walking Together for Growth" scenario. The degree of optimism, expressed as a percentage increase in growth between 2011 and 2030, amounted to 10% for population, 20% for employment in manufacturing and trade and 10% for employment in most other sectors. Some of this optimism has since materialized in the form of foreign direct investment in motor vehicle manufacturing.
3. Two other scenarios were simulated which only differed in respect of which urban node would be prioritized for investment in bulk-infrastructure and catalytic projects to attract external investment, Coega (marker 1 in Figure 9) or Jachtvlakte (marker 2 in Figure 9). Both scenarios were based on the optimistic demographic and employment growth projections.
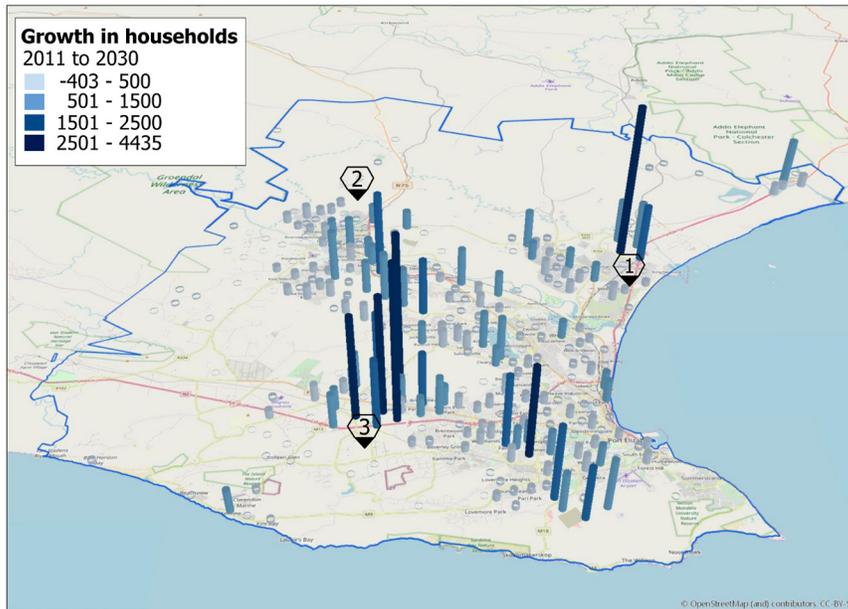
7.3.3   Results
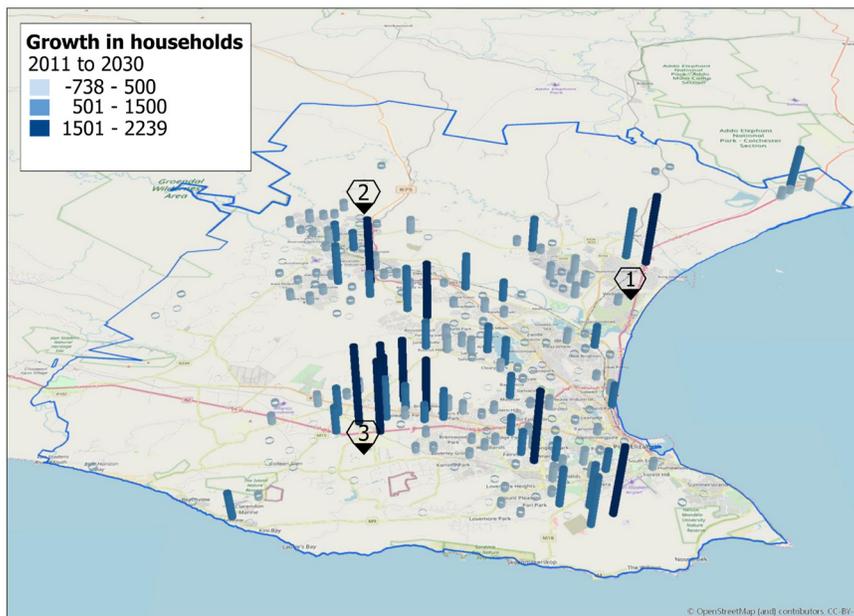
Figure 9. Trend scenario; the height of each bar represents the growth in number of households per analysis zone between 2011 and 2030
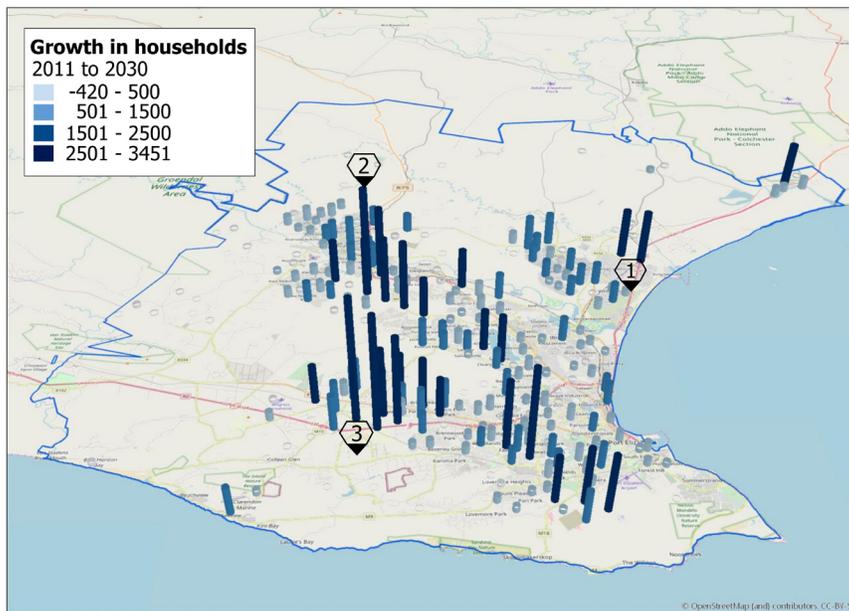

Figure 10. Trend scenario with optimistic growth; the height of each bar represents the growth in number of households per analysis zone between 2011 and 2030

The spatial distribution of growth is not proportional to the aggregate growth for the city. This is evident from a comparison between Figures 9 and 10, where the higher growth projections resulted in a visibly different distribution between nodes. This is simply a function of what residential and non-

residential land is available for green-field and infill development. This not only confirms the value of urban simulation to quantify growth in spatially explicit terms but hints at the considerable influence that municipalities have on future urban form through the strategic release of land and the provision of bulk infrastructure.

Figure 11. Coega priority node; the height of each bar represents the growth in number of households per analysis zone between 2011 and 2030
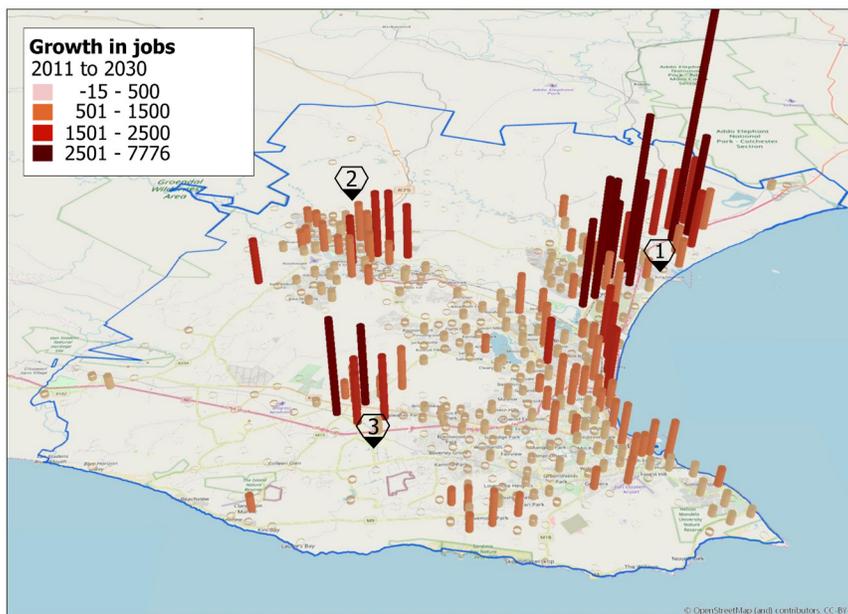
Figure 12. Coega priority node; the height of each bar represents the growth in number of jobs per analysis zone between 2011 and 2030

Although one would expect a take-up of residential development near employment opportunities, especially by less affluent households wanting to spend as little as possible on transportation to work, this is not always the case. A comparison between Figures 11 and 12 shows that while the industrial development in the Coega node (marker 1) did indeed result in a take-up of residential development in the area, there was also take-up in the N2 node (marker 3). This might be an unintended consequence of the comprehensive transportation system that the model was based on, in the sense that it enables more households to commute over greater distances and thus greater choice in deciding where to live and where to work. Given such a choice, many households (especially those with higher incomes) could well prefer not to live close to industrial areas. The N2 node is favored by property developers because of its accessibility to the N2 freeway and clean industry character. There was some confirmation of this in the observation that the average household income for zones that grew in the N2 node was indeed higher than those in the Coega node.

8 Conclusions

Given that this work represents the first fully interactive land use transportation modelling done in South Africa, Urbansim-SA has by no means been institutionalized in the sense of being formally adopted by the cities to support the planning processes alluded to in Section 2. The objective was to take some early adopters in each of the cities through the simulation process to prove the concept. Rather than using the model in the preparation of MSDF, CIF & BEPP as suggested in the introduction, various aspects of plans already prepared without the model were investigated/validated.

A surprising finding of the literature review of UrbanSim applications in Europe/Asia was that all except one used zone/grid cell geometries. For the work reported herein we used the parcel geometry from the outset to avoid behavioral aggregation as discussed in Section 3.2 and because it provided us with a mechanism to model backyard dwellings by combining formal and informal template components into a new template for the parcel. Another factor that contributed to this approach was that the drivers/sponsors of the case study projects were invariably urban planners (known as Town and Regional Planners in South Africa) with a live interest in the classification and consumption of land at parcel level.

The results obtained from all three case studies confirmed that policy decisions can be influenced by using ILUT modelling to add elements of rational analysis and hypothesis testing to a controversial decision-making process. In the City of Tshwane case study, "rational analysis" was for example introduced simply by comparing the increase in density (hu/ha) within the densification zones proposed by the Trend and Priority Areas scenarios over the simulation period. The results clearly showed that the Priority Areas scenario had a much better chance of increasing public transport ridership levels to contain subsidies. This, together with the projected savings in the cost of providing bulk waste-water infrastructure, provided invaluable support to proponents of the Priority Areas scenario. The Ekurhuleni case study probably provides the best example of hypothesis testing where different institutional stakeholders held opposing views about the best location for government funded housing projects. Some supported large projects, motivated by the economies of scale that could be achieved in management and construction, even though such projects would require tracts of land, which would most likely only be found on the fringes. Another department promoted smaller infill projects as close as possible to employment opportunities. By considering the two views as different hypotheses, the modelling concluded that the expected benefit to the 100 000 affected households would not nearly be as significant as one might intuitively think. The fact that this could be attributed to the availability of subsidized public transport (especially rail) in proximity to the large projects not only demonstrated cause and effect to the

participants but also surfaced preconditions for the success of policy decisions. For example, if the decision went the way of large projects but commuters were to lose the benefit of subsidized transport, for example due to concerns about personal safety on trains, history would show that infill projects might have been the better policy decision.

Based on the sample results from the three case studies as well as the results that were not presented here, cities in South Africa were found to be unique from an ILUT modelling point of view. The model specification for one city (including the selection of explanatory variables) will simply not work in another city even though the context appears to be similar. Calculated densities for development templates of the same Knowledge Factory typology differed substantially and new sub classes even had to be created in some cities to resolve anomalies. What did, however, consistently stand out is the considerable influence that municipalities have on the future urban form through the strategic release of land and provision of bulk infrastructure. If this is used as a tool to guide development, municipalities should guard against creating artificial shortages that will inflate the price of land, except if it serves a valid policy goal and land for lower-income housing can be provided by different means, for example state-owned land.

While it has been possible to adapt UrbanSim to succeed in South Africa, the applicability of the work to other developing countries will depend entirely on the availability of a good census (preferably two) and finding substitutes for the Knowledge Factory and Geoterra Image datasets, without which the work in South Africa would not have been possible. If this data can be collected, the South African journey has shown that ILUT modelling adds an element of rational analysis and hypothesis testing that can contribute to policy formulation and analysis.

9   Data

Three of the data sources used in our journey are subject to license agreements. These include the Knowledge Factory Cluster Plus product, the Geoterra Image Building-Based Land-Use product and the IHS Markit Population and Employment projections. The details of the three companies are provided in the references to assist researchers that may want to reproduce our results.

Acknowledgements

References

Africa.com. (2018). Home ownership in South Africa. Retrieved from https://africa.com/home-ownership-in-south-africa/

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

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

Cervero, R. (2013). Linking urban transport and land use in developing countries. *Journal of Transport and Land Use, 6*(1), 7–24.

Department of Human Settlements. (2019). Delivery of serviced sites and houses/units from the HSDG since 1994 (20 years). Retrieved from http://www.dhs.gov.za/sites/default/files/documents/statistics/20 Year delivery Sites & Houses HSDG finalised ver. 29052014.pdf

Developing Countries Population. (2019). Retrieved from http://worldpopulationreview.com/countries/developing-countries/

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.

Felsenstein, D., Axhausen, K., & Waddell, P. (2010). Land use-transportation modeling with Urban-Sim: Experiences and progress. *Journal of Transport and Land Use, 3*(2), 1–3.

Felsenstein, D., Lichter, M., & Ashbel, E. (2014). Coastal congestion: Simulating port expansion and land use change under zero-sum conditions. *Ocean & Coastal Management, 101,* 89–101.

Geofabrik GmbH. (2018). OpenStreetMap data for South Africa. Retrieved from http://download.geofabrik.de/africa/south-africa-and-lesotho.html

Geoterra Image. (2001, 2011). Building based land use. Proprietary vector data set. Retrieved from http://www.geoterraimage.com/pdfs/101 Building Based Land Use.pdf

IHS Markit. (2012, 2016). Proprietary city-wide demographic and employment projections. Retrieved from http://www.ihsmarkit.co.za

Jin, J., & Lee, H. (2018). Understanding residential location choices: An application of the UrbanSim residential location model on Suwon, Korea. *International Journal of Urban Sciences, 22*(2), 216–235.

Joo, Y., Mehedy Hassan, M., & Jun, C. (2011). An application of the UrbanSim land price model in Yongsan-gu, Seoul, Korea. *International Journal of Urban Sciences, 15*(1), 15–24.

Knowledge Factory. (2006). Discontinued proprietary micro-marketing data set. Company profile retrieved from http://www.knowledgefactory.co.za/

Kryvobokov, M., Mercier, A., Bonnafous, A., & Bouf, D. (2015). Urban simulation with alternative road pricing scenarios. *Case Studies on Transport Policy, 3,* 196–205.

Lechner, T., Watson, B., Tisue, S., Wilensky, U., & Felsen, M. (2004). *Procedural modeling of land use in cities* (Technical report NWU-CS-04-38). Evanston, IL: Northwestern University, Center for Connected Learning and Computer-Based Modeling.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). New York: Academic Press.

Moeckel, R. (2017). Constraints in household relocation: Modeling land-use/transport interactions that respect time and monetary budgets. *Journal of Transport and Land Use, 10*(2), 1–18.

National Treasury. (2015). Cities support program: Guidance note for the built environment performance plan. Retrieved from https://csp.treasury.gov.za/Projectdocuments/BEPP Guidelines 2016_17.pdf

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.

Picard, N., de Palma, A., & Kiarash, M. (2015). Application of UrbanSim in Paris (Ile-de-France) case study. In M. Bierlaire, A. de Palma, R. Hurtubia & P. Waddell (Eds.), *Integrated transport and land use modeling for sustainable cities* (Chapter 20). Abingdon, UK, and Lausanne, Switzerland: Routledge and EPFL Press.

Rust, K. (2012, October). *Opportunities in South Africa's housing finance & delivery framework: Navigating the Gap.* Paper presented at the Affordable Housing Indaba, Gauteng, South Africa.

Schirmer, P., Zöllig, C., Müller, K., Bodenmann, B., & Axhausen, K. (2011, September). *The Zurich case study of UrbanSim.* Paper presented at the 51st ERSA Conference, Barcelona, Spain.

Shi, J., Tong, X., Zhang, H., & Tao, D. (2013). Spatial interaction of urban residence and workplace: An UrbanSim application in Yichang, China. *Beijing Daxue Xuebao (Ziran Kexue Ban)/Acta Scientiarum Naturalium Universitatis Pekinensis, 49,* 1065–1074.

South African History Online. (2016). A history of Apartheid in South Africa. Retrieved from http://www.sahistory.org.za/article/history-apartheid-south-africa

Stats SA. (2013). National household travel survey. Retrieved from http://www.statssa.gov.za/publications/P0320/P03202013.pdf

Stats SA. (2015). Measuring household expenditure on public transport. Retrieved from http://www.statssa.gov.za/?p=5943

Stats SA. (2018). Victims of crime survey 2017/18. Retrieved from http://www.statssa.gov.za/?page_id=1854&PPN=P0341&SCH=7373

Stats SA. (2019a). Census 2011 statistical release. Retrieved from https://www.statssa.gov.za/publications/P03014/P030142011.pdf

Stats SA. (2019b). General household survey 2018. Retrieved from http://www.statssa.gov.za/?page_id=1854&PPN=P0318&SCH=7652

Stats SA. (2019c). Quarterly labor force survey, Quarter 2: 2019. Retrieved from http://www.statssa.gov.za/?page_id=1854&PPN=P0211&SCH=7620

Vanegas, C., Aliaga, D., Müller, P., Waddell, P., Watson, B., & Wonka, P. (2009). Modeling the appearance and behavior of urban spaces. *Computer Graphics Forum, 29*(1), 25–42.

Waddell, P. (2002). UrbanSim: Modeling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association, 68*(3), 297–314.

Waddell, P. (2011). Integrated land use and transportation planning and modeling: Addressing challenges in research and practice. *Transport Reviews, 31*(2), 209–229.

Waldeck, L., & van Heerden, Q. (2017). Integrated land-use and transportation modelling in developing countries: Using OpenTripPlanner to determine lowest-cost commute trips. In I.M. Schoeman (Ed.), *Transportation, land use and integration: Applications in developing countries.* Southampton, UK: WIT Press.

Wegener, M. (2007). Themes and relationships. In S. Marshall & D. Banister (Eds.), *Land use and transport.* Oxford, UK: Elsevier.

World Bank. (2011). GINI index. Retrieved from https://data.worldbank.org/indicator/SI.POV.GINI

Ye, X., Konduri, K., Pendyala, R., Sana, B., & Waddell, P. (2009, January). *A methodology to match distributions of both household and person attributes in the generation of synthetic populations.* Paper presented at the 88th Annual Meetings of the Transportation Research Board, Washington, DC.

Ziemke, D., Nagel, K., & Moeckel, R. (2016). Towards an agent-based, integrated land-use transport modeling system. *Procedia Computer Science, 83,* 958–963.