ETH zürich

ETH Library

Towards comparability in residential location choice modeling

A review of literature

Working Paper

Author(s):
Schirmer, Patrick M.; van Eggermond, Michael A.B.; Axhausen, Kay W.

Publication date:
2013

Permanent link:
https://doi.org/10.3929/ethz-b-000441314

Rights / license:
Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 962

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

THE JOURNAL OF TRANSPORT AND LAND USE http://jtlu.org
VOL. X NO. X [201x] pp. 1–20 doi: 10.5198/jtlu.vxix.nnn

Towards comparability in residential location choice modeling - a review of literature

Patrick M. Schirmer
IVT ETH Zurich [a]

Michael A.B. van Eggermond
Future Cities Laboratory, Singapore ETH Centre

Kay W. Axhausen
IVT ETH Zurich

Abstract: With the availability of disaggregated spatial data, residential location choice models have become more and more diverse in the hypothesis tested. This reduces the comparability of the main findings in different studies, although one might expect a comparable human behavior for comparable study areas. In the paper we review recent literature of the residential location choice models on a disaggregated level, in order to conclude in a set of variables recommended for use in models and simulations.

Keywords: residential location choice; agent-based modeling; urban simulation; planning support systems; discrete choice model

*Todo list

1 Introduction

The residential location of households is one of the key elements for urban dynamics. Amongst others, it has an impact on employment conditions, economic development, social structure, segregation and the transport system. Understanding the residential location choice behavior has thus been a primary concern to urban planners, policy-makers and researchers.

The roots of residential location modeling can be traced back to the first advances in land-use modeling by von Thunen (1826). Von Thuenen made the first attempt to explain the effect of transport costs on the location of activities and the functioning of the land market by means of a single-market in an agricultural region where land-owners are willing to rent their properties to the highest bidder: the bid-rent concept. Alonso (1964) applies this bid-rent concept to residential location and considers a mono-centric city with employment opportunities. Individuals and households base their residential location choice of households on maximizing a utility function that depends on the expenditure in goods, size of the land lot, and distance to the city center.

Parallel to Alonso, Lowry (1964) applies the gravity model for residential location. Lowry assumes an initial set of basic employment centers per zone. Households are allocated to zones based on a deterrence function describing the number of workers employed in zone and living in zone. Residential attractiveness is measured by the amount of land available for development in a particular zone.

The disaggregated modeling approach of the discrete modeling framework propagated the development of land use simulations, in which a series of location choice models interact (Wegener 2004). With the increasing availability of disaggregated data, agent based land use and transport simulations have gained widespread attentionand planning support systems for decisions in planning policy, e.g. in the Puget Sound Area and in the San Francisco Bay Area UrbanSim has been applied, in ILUTE is implemented for Toronto and MATSim has been uses to evaluate different transport scenarios. Com-

[a] patrick.schirmer@ivt.baug.ethz.ch

Copyright 201x IVT ETH Zurich.
Licensed under the Creative Commons Attribution – NonCommercial License 3.0.

prehensive reviews can be found in Wegener (2004) and Iacono *et al.* (2008).

With the increasing availability of highly disaggregated data, the possibilities of defining the residential alternative and its attributes have changed. Whereas earlier studies used administrative districts or transport zones as choice alternatives (Anas 1982; Weisbrod *et al.* 1980), and aggregated characteristics of these zones as attributes, recent models consider buildings or units as choice alternative and include building specific attributes as well as location specific attributes (Habib and Miller 2009; Lee and Waddell 2010a).

The range of hypothesis and attributes that have been tested within residential location choice models has varied, which led to a reduced comparability of the models. Moreover the diversity of attributes that is used in residential location choice prevents the definition of a common data model for simulation of choices processes in different study areas (Figure ??). Finally, the question arises if all variables included in residential location choice models can be used in a land-use simulation.

In order to enhance the usability of disaggregated land use and transport simulations for planning & decision support systems, future work will benefit by using a common data model and generalized model specifications, at least as a starting point for the setup of a land-use simulation. In order to achieve this goal, this papers aims to investigate systematically which factors influence residential location choice by means of a comprehensive literature review. The purpose of the literature review is to classify used variables according to their use in simulation process, find common attributes between studies that proved to be relevant, understand the various interaction terms and summarize the main findings in literature. This review can be found in section 3. An initial proposal for a common data model can be found in section 4. The next section continues with discrete choice models and their application to residential location choice.

2 Modelling residential location choice

The discrete choice framework was first introduced to residential location choice by McFadden (1978). Initial studies considered households which move to a certain zone (Anas 1982; Weisbrod *et al.* 1980). In recent years the parcel and residential unit have been considered as choice alternative (e.g. Lee and Waddell 2010a). This section will discuss discrete choice models and their application to residential location choice and two specific aspects of discrete choice models: the utility function and choice set formation. Section 3 will continue with specific attributes considered in a wide range of residential location studies.

2.1 Discrete Choice Models

Within the discrete choice framework, a decision-maker chooses from a set of alternatives. Each alternative is assumed to have a number of attributes. Each attribute has a level of utility or disutility, which capture the costs and benefits of an alternative; the utility U of an alternative i for a decision-maker q is defined by $U_{iq} = V_{iq} + \varepsilon_{iq} = f(\beta_i x_{iq} + \varepsilon_{iq})$ with a deterministic part V_{iq} that consists of a function f of the vector β_i of taste parameters and the vector x_{iq} of attributes of the alternative, the decision-maker and the choice situation. In addition, socio-demographic attributes of decision-maker q can be included in the deterministic part of the utility function. The non-deterministic, non-observable part of the utility function is captured by ε_{iq}.

Decision-maker q will choose the alternative from set C with the highest utility $P(i|C_q) = P[U_{iq} \geq U_{jq} \forall j \in C_q] = P[U_{iq} \max_{j \in C_q} U_{jq}]$.

The most commonly used discrete choice model is the Multinomial Logit (MNL) Model due to its ease of estimation and simple mathematical structure (McFadden 1974). It is based on the assumption that the random terms, often called error terms or disturbances, are identically and independently (i.i.d.) Gumbel distributed. The Independence of Irrelevant Alternatives (IIA) property states that the ratio of the choice probabilities of any two alternatives is entirely unaffected by the systematic utilities of any other alternatives (Ben-Akiva and Lerman 1985). This property stems from the fact that the distribution of the disturbances are assumed to be mutually independent and requires that the sources of errors contributing to the disturbances do so in a way that the total disturbances are independent. The IIA property can be overcome by applying models that allow for a non-zero covariance matrix or use a nested choice structure. An example of the prior are Mixed Logit models, an example of the latter Nested Logit and Cross-Nested Logit model.

2.2 Discrete choice models in residential location choice

Approaches

The type of model applied depends on the choice dimensions considered (i.e. car ownership, residential mobility, residential location). The majority of studies solely considering residential location choice employ the MNL model (e.g. Guo and Bhat 2002; Lee and Waddell 2010a; Vyvere *et al.* 1998). Other studies apply Mixed Logit models to account for heterogeneity amongst decision-makers (e.g. Eluru *et al.* 2009; Habib and Miller 2009; Zhou and Kockelman 2008). Nested Logit models are applied when multiple dimension are considered; each dimension is than considered to be a seperate nest. Examples are residential relocation and location choice (Andrew and Meen 2006; Eluru *et al.* 2009; Lee and Waddell 2010b), residential location and activity pattern (Ben-Akiva and Bowman 1998; Eliasson 2010) and residential location and auto ownership (Weisbrod *et al.* 1980). Zondag and Pieters (2005) consider the move or stay decision, followed by a nest representing the decision to remain the current region or move to a different region; they find that moving is a local process.

Utility function

The utility function of each alternative contains a number of attributes. The relative weight of the parameter estimates for these attributes give insights in the trade-offs decision-makers carry out, such as the trade-off between alternative specific attributes (location) and socio-demographics (income, age, household composition).

Most studies use cross-sectional data to model residential location; the previous location of the household and changes in household demographics are not taken into account. However, evidence can be found that the previous location plays a role in choosing a new location - households who preferred shorter commute distances at their previous location do so at their new location (Chen *et al.* 2008). Households also prefer a gain in bedrooms and gains in open area (Habib and Miller 2009), which may be led by a change in household composition.

Cascetta *et al.* (2007) propose a methodology to identify dominance attributes which may be defined in different ways, in accordance with the specific choice context, and in which way they can be introduced as perception attributes in random utility models. Their estimation results show a generally high significance of all these attributes and a considerable improvement in the model's goodness-of-fit statistics.

2.3 Choice set formation & generation

Delete if too many words: Every choice is made from the set of the alternatives or the choice set several approaches are discussed in literature to determine the choice set which contains the alternatives that were available to the decision maker as this universal set of alternatives is unknown to the analyst in a revealed preference environment. (Swait 2001) proposes to formulate several choice sets (a set of choice sets) and estimate the probability of a choice set being the true choice set. On the other hand an attempt can be made by following heuristics considered by the decision-maker and thus acknowledging that choice set formation is a dynamic search process. Another way to avoid the burden of working with extremely large choice sets is to estimate parameters from a subset from alternatives in the MNL model where the independence of irrelevant alternatives (IIA) is assumed, as demonstrated by (McFadden 1978).

Most studies consider the universal choice set or sample from the universal choice set (e.g Chen *et al.* 2008; de Palma *et al.* 2007; Guo and Bhat 2007; Habib and Miller 2009; Lee and Waddell 2010a,b). de Palma *et al.* (2007) compare different sample sizes. Weighting of the sampling is done according to the number of units per zone. Their findings confirm that variances decrease proportionally to sample size, and that significant estimated parameters remain constant. Insignificant estimated parameters, such as number of stations, travel time to highway, however vary in sign and size. The second approach, considering heuristics followed by the decision-maker, has been considered by Zolfaghari *et al.* (2012). They apply a hazard based model and set thresholds on acceptable property price and commuting times. However, they find that their choice set formation is outperformed by a choice set formed by random sampling. (Cascetta *et al.* 2007) include the aforementioned dominance variables in the sampling of alternatives. Their results show that the weighted sampling gives parameters' estimates 'closer' to those obtained with full choice set.

3 Measuring residential location choice

In this section we explore dependent variables that have been reported in previous studies on residential location choice models and summarize the findings of these studies. Commonly a distinction is made between household attributes, attributes of the residence and location attributes. Additionally we will classify location attributes depending on the type of data they are derive of and the calculation method they use.

Estimates of different models do not allow for direct comparison if not using standardized variables or elasticities. Also within a single model the estimates cannot be compared due to different scales of the data source. The comparison of elasticities, e.g. to a price variable would allow for comparison, but unfortunately is not being done in most studies. Within this review we thus can only report the level of significance and the positive or negative influence on the resulting utility of a location, as well as summarize the main findings of the authors.

As figure 1 shows, the models in the reviewed literature are very divers and use a wide range of variables. Each of the variables might capture some correlation effects to unused variables, which again will be different between the reported models. If these correlated variable are used in another model, it will result in different estimation results. This creates difficulties when trying to discuss on models that have been reported. Nonetheless we need to explore and group the variety of variables that have been used in previous literature to search for common observations, that captures common human behaviour.

3.1 Household attributes

Variables describing the household, are commonly included in residential location models. These socio-economic characteristics are interacted with other variables in the utility function in order to estimate taste preferences across different household segments. An important driver in residential location choice is household income, which is often interacted with rent or mortgage payments. This will be discussed in the section 3.2.

Distance to previous location and social networks

Only few studies explore the influence of distance to previous location on the utility of a new residential location. Axhausen *et al.* (2004) and de Palma *et al.* (2005, 2007) find that households tend to stay close to their previous location or remain in the same district. Zondag and Pieters (2005) combine the Euclidean distance in its logarithmic transformation between previous and current location and find it to be the most dominant variable in their model for various household types tested, all being negative except for households over age 65.

Gordan (1992) mentions the desire to maintain social networks; it can thus be expected that this variable is of relevance for residential relocation. Vyvere *et al.* (1998) introduce the distance to social contacts in their model specification and show a negative impact on residential utility. Belart (2011) further explores this effect and observes a preference for proximity to social contacts when using distance to social contacts weighted with by number of meetings per month. As shown in the complementary modeling study of this paper (Schirmer *et al.* 2013) this variable correlates with the previous location: it is unclear whether location is influenced by social contacts or whether the social network is formed by the location.

Life cycles and lifestyles Various studies have shown the relocation probability depends on life cycle events such as the change of marital status, job change, starting a family, size of household, retirement and age (Andrew and Meen 2006; Eluru *et al.* 2009; Kim *et al.* 2005; Lee and Waddell 2010a). In early stages of the life-cycle households tend to move close to the city center where as in later stages households tend to move away. This observation is also captured by the different preferences for urban characteristics (see chapters ??, 3.3). Young individuals tend to prefer mixed use areas with a higher population density, while others household types prefer areas with a lower population density.

As distance to previous location is disliked by relocating households we would assume that households tend to stay on the same location, which is contradictory to the observation of life cycle relocations.

Beside the differentiation of households based on their life cycles, several studies will use the concept of lifestyles instead. Lifestyles are an approach of the social sciences that have been explored since the 1980s and have proven to be of explanatory value for transportation models. Persons are grouped based on their common behaviour in daily life and their cultural, social and leisure behaviour (Müller 1992). Belart (2011) explores self-selection-effects as alternative to the life cycle approach, based on the classification of Otte (2005) and an own classification. He differentiates four groups of lifestyles which slightly enhance model fit when being interacted

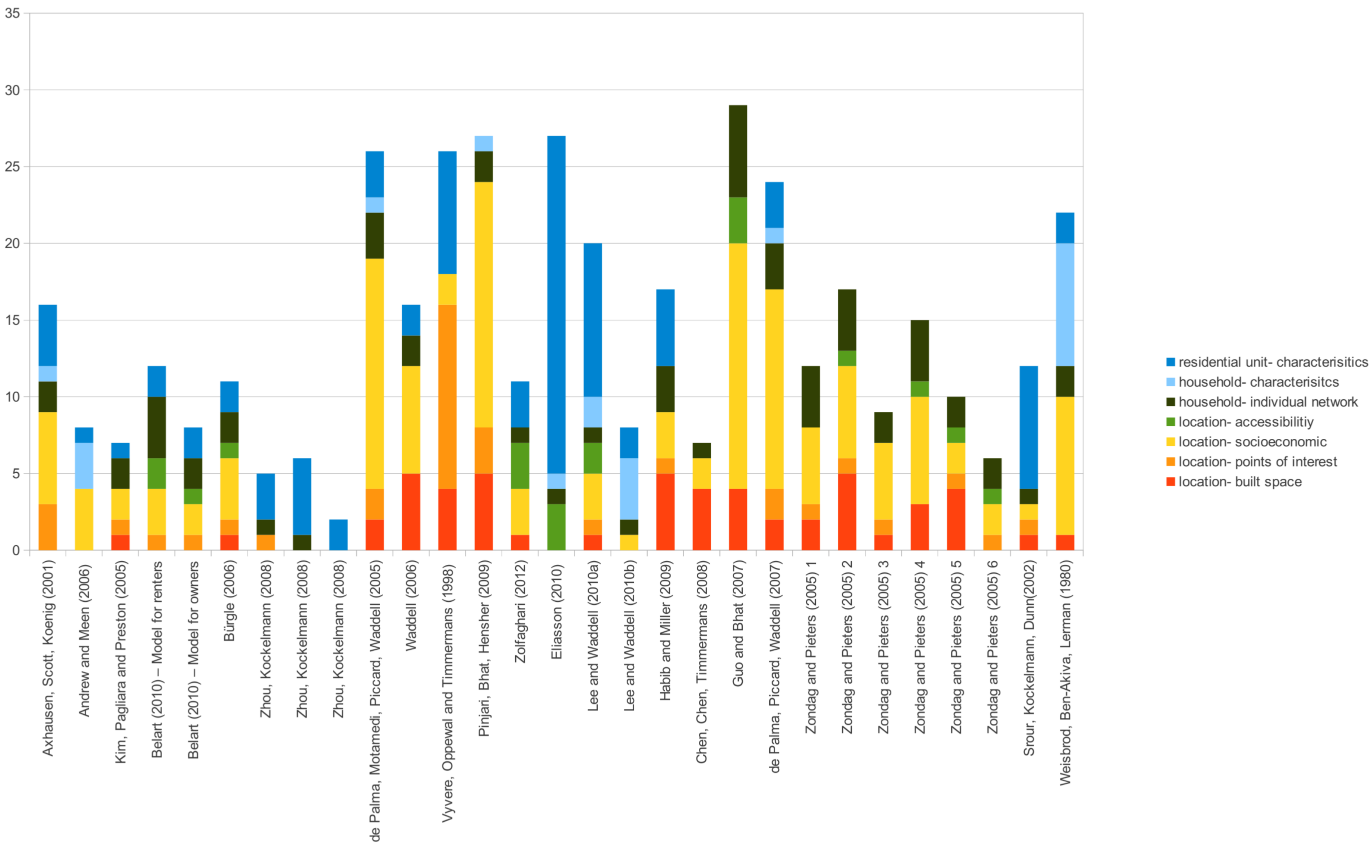

Figure 1: Amount of variables reported for residential location choice

with population density, open space, travel time and percentage of similar household in neighbourhood.

Walker and Li (2007) differentiate between three types of lifestyles that reveal different preference for urban density, retail and service density as well as lot sizes, with car oriented households having a preference for lower densities and bigger lot sizes. (Krizek and Waddell 2003) differentiate between nine types of lifestyles that show differences in location choice behaviour and mobility behaviour as well by interaction with urban density and travel distances.

However, to the knowledge of the author a systematic comparison between an approach considering lifestyles versus one considering life cycles and other characteristics of the person as educational background, income or ethnic background, is still outstanding. The variation in the lifestyle clusters found in literature does not allow a comparison of the model results which reduces the model explanation; the evaluation of the lifestyle-concept versus the alternative attributes thus demands further research.

3.2 Residential unit

Choice alternatives within residential location alternatives have different scales for instance due to scope of the study or data limitations. A common approach is to use a zone as residential alternative (e.g. Axhausen *et al.* 2004; Chen *et al.* 2008; de Palma *et al.* 2007; Guo and Bhat 2007; Pinjari *et al.* 2011), while few recent studies consider the building as alternative (e.g. Habib and Miller 2009; Lee and Waddell 2010a,b; Vyvere *et al.* 1998). If the residential unit is considered as choice alternative, it is possible to interact household attributes with attributes of the residential unit, e.g. size, price and number of rooms. Lee and Waddell (2010a) observe that these dwelling characteristics tend to dominate over accessibility indicators. Due to the different scales, also the variables that can be used to describe alternatives are divers. Zonal attributes are mainly captured by the location attributes, to be discussed in section 3.3. In this section we will focus on disaggregated characteristics of the residential unit in form of parcels, buildings and dwellings.

Costs, price and value

The sales price or rental costs are often included in the model specification in the reviewed studies as price is an important determinant in choosing a residence. Price captures various location characteristics as can be seen in various hedonic regression models (e.g Löchl and Axhausen 2010), but nonetheless all models that include property price mention its significant negative impact. While some studies implement price as untransformed value (Andrew and Meen 2006; Kim *et al.* 2005; Vyvere *et al.* 1998; Zolfaghari *et al.* 2012) other studies include a logarithmic transformation (de Palma *et al.* 2005, 2007; Habib and Miller 2009; Lee and Waddell 2010a,b). Additionally several studies include a ratio of price with the households income (Habib and Miller 2009; Zolfaghari *et al.* 2012). Zolfaghari *et al.* (2012) observes a postive impact of price when constraining the choice set based on price, which can indicate households choosing the best available alternative within a certain range. The ratio of price to household income is also of significance when it is the only price related variable in the model specification (Bürgle 2006; Belart 2011; Lee and Waddell 2010a; Waddell 2006; Weisbrod *et al.* 1980; Zhou and Kockelman 2008). Walker and Li (2007) observe that price-sensitivity reduces with rising income; the integration of the ratio and eventually the logarithmic transformation seems reasonable. Alternatively interaction between price and income groups can further improve model estimates. As Axhausen *et al.* (2004) state that owners and tenants of shared accomodation have a higher willingness to pay for their location compared to other users, a further differentiation of these groups is recommended. Srour *et al.* (2002) and Waddell (2006) also explore the effect of the improvement value of a residence. Both studies report that this variable is valued positive by all households types tested, which is not surprising as a high value of a building, representing high quality can be expected to be favoured by households. Nonetheless it should be mentioned that Weisbrod *et al.* (1980) did not find any significant impact on this variable. One important issue that arises with inclusion of price in residential model is the issue of price endogeneity: the unit's price is correlated with the model;s error term. This can be because variables are omitted that are correlated with price. Guevara (2005) addresses this issue and introduces a two-step estimation method to overcome this source of endogeneity.

Unit size

Size of the residential unit can be included as absolute value (Zhou and Kockelman 2008) or as a ratio using the household size to capture the space per person (Bürgle 2006; Belart 2011). Axhausen *et al.* (2004) subtract the observed mean space per person from the individual space per person to capture regional differences. These approaches yield similar results: households prefer to have more space per person when relocating.

In addition to dwelling size the number of (bed)rooms has been considered in several studies. Habib and Miller (2009) observe a positive impact for a gain in number of rooms in their reference dependent model. Eliasson (2010) present an approach similar to a two-step regression to measure the number of rooms in reference to the floor space. This formulation allows implementing both variables and avoids multicollinearity between floor space and number of bedrooms. Their model estimates show that single-person households prefer a lower number of rooms with more floor space per rooms, while all non-single households favour additional rooms instead. The integration of number of rooms into the residential choice model can thus be expected to enhance the explanatory power of the model, but eventually demands for correction of correlation with the floor space. Similar to the floor space variable, the number of rooms might then be integrated in relation to household size, although no study reported this approach.

Housing typology

Various studies have observed the preference of a specific housing typology. However due to differences in the historic heritage of cities and different structures within the residential real estate market, such as differences between public and private housing, tenant protection and mortgages types different preference for different areas are to be expected. Vyvere *et al.* (1998) finds a general preference for houses in Belgium; Habib and Miller (2009) finds a negative preference for attached houses in Toronto. Lee and Waddell (2010a,b) differentiate between household types and find single person households and renters to prefer for multifamily houses; households with children favour single family buildings. Axhausen *et al.* (2004) observes that households prefer a similar type of housing as compared to their previous dwelling and proposes to include the previous location type as variable. Eventually this can capture self-selection effects; households with different lifestyles prefer a preference for certain housing typology. On the other side this would ignore changes in household composition that might be the reason to relocate. In conclusion housing typology is expected to be of relevance for residential location choice and differs per household type: families favour single family (detached) houses and single person households being more attracted by multi-family houses.

Dwelling features

Vyvere *et al.* (1998) finds the number of garages to be of importance to households when relocating, although we can expect that this is only of importance to households owning a car. In the same study it is shown that households dislike buildings being constructed before 1960. The value of historic buildings has not been explored explicitly, although Srour *et al.* (2002) find the average age of buildings in a zone to be valued positive.

3.3 Location attributes

The manner in which attributes describing the residential location are included in the choice alternatives differs in the different studies reviewed. First, the detail of the location attributes depends on the resolution of the choice set alternative (e.g. zone, neighbourhood, unit) and the disaggregation the data available for the study. For instance, zonal information will not allow the exploration of local neighbourhood effects. Guo and Bhat (2007) explore the perception of the neighbourhood using different spatial extents and calculation methods. They differentiate between fixed neighbourhood boundaries, based on census tracts and blocks, against sliding neighbourhood extents, using Euclidean distances and network distances, and show that the calculation has influences parameter estimates. Their findings also show that the neighbourhood effects differ between attributes describing the built environment (structural) and the socio-economic distribution, being smaller for the latter.

In addition to the level of aggregation the spatial representation of data varies. First, data concerning the socio-economic distribution is located in space either as a abstract grid cell, a polygon or as a spatial point; these spatial forms contain data being the result of an aggregation of individual data points. Second, cadastral data such as buildings, parcels and blocks possesses a geometric or volumetric shape which occupies space and interacts with other shapes. A specific form of cadastral information describes the locations of public functions and ignores the spatial representation by summarizing attributes to a spatial point, e.g. the location of a hospital or school. The last form of spatial data comprises of the network of roads and public transport, represented through abstraction in form of spatial lines, which connect locations in a city .

Software for urban simulation (e.g. Wegener 2004; Zöllig *et al.* 2011) generally cannot handle all forms of spatial data and mainly considers the socio-economic distribution in the simulation process. This data structure reduces the options to simulate decisions processes including spatial information within the simulation (see section 4). As the purpose of this paper is to explore the types of variables used in residential location choice in order to understand their impact and usage in urban simulation processes we will address this current limitation and differentiate location attributes into attributes representing the built environment (shapes), networks, points of interest and the socio-economic environment.

Location attributes are characterized by their spatial distribution and require a variety of geographic functions as part of the data preparation. Their use in the simulation process of integrated land use simulation might be limited as it depends on the data model and the methods implemented. As not all of those methods are implemented in urban simulation software, we will further distinguish these processing techniques and group the reviewed variables based on:

- Buffer
- Distance (Euclidean distance, network distance, travel time, travel cost)
- Density (kernel density, number of object in spatial reference)
- Spatial ratio (share, fraction of area in spatial reference, fraction of population)
- Accessibility (weighted distance function)

Built environment

The *built environment* is represented by the geometries and volumes of spatial objects around a chosen residential unit and includes buildings, parcels, blocks and the connecting networks. This includes buildings, parcels, blocks, extents of urban zones and the connecting network, and can only be changed by construction activity.

As most location choice models are based on census data the use of built space variables derived of cadastral information is not very common though. Within the reviewed literature we can differentiate variables on landuse mix, open-space, structural densities, built densities, network buffers and settlement areas.

Built density

Waddell (2006) finds dwelling density to have a negative impact on the residential utility for all family households in the Puget Sound Region. The density of dwelling units as logarithmic expression however has a significant and positive influence for all household types tested.

Similar to results found when using population density (section), the dwelling density has a positive sign for young households. This is not surprising as it is expected to be correlated with population density. Built space information can be used as alternative to certain socio-demographic variables when not given.

Structural density

Pinjari *et al.* (2009) includes structural variables based on the building shapes and urban form and introduce the length of networks and number of blocks per square mile. The number of bike lanes is preferred by all households whereas the number of blocks has a negative impact for households with a high income. This is the only study to include such kind of variables. Nevertheless, these measurements form an interesting approach to represent geometric information and urban characteristics and demand for further research.

Network and noise

Bürgle (2006) finds that the proximity to major roads or railways has a negative effect on residential utility in the Zurich area. This can be interpreted as an indicator to noise which also is observed by Vyvere *et al.* (1998) within their stated preference survey in Louvain-la-Neuve (Belgium). A positive influence is found by Waddell (2006) for Seattle and de Palma *et al.* (2005) in Paris who observe a positive impact of proximity to arterials and highways. In de Palma *et al.* (2005) the study is on an aggregated zonal level and both studies do not include accessibility calculations. The distance to highway can then be expected to capture accessibility effects instead of noise effects.

Open space

The use of fraction of open space and green area as variables in residential location choice has been reported by several authors, although it is often undefined whether the authors mean recreational areas and/ or unbuilt area. Habib and Miller (2009) find the percentage of green area in the neighbourhood to have a positive impact for households in Toronto. Similar results have been found by Chen *et al.* (2008) in Seattle and - for open space - also by Zondag and Pieters (2005) in the Netherlands, with both studies interacting the variable with various household types. The latter study additionally finds a positive impact of percentage of water surface on the residential utility for workers, which reflects the architectural attraction toward waterfronts on the residential housing market in the Netherlands. Guo and Bhat (2007) finds a negative impact of open space variables for couple-only households in San Fransisco. As these results are specific to only one household type, we expect open space to have a positive impact for all different life-cycles as well as household types.

Land-use

Land-use mix has been explored in residential location models; the exact specification however is not discussed in detail in most studies. A common observation is the negative impact of industrial land use in proximity of the residential location (Habib and Miller 2009; Weisbrod *et al.* 1980); mixed land-use therefore often only differentiates between residential and commercial (office and retail) rather than including industrial land-use. Waddell (2006) includes an indicator for mixed land-use in his study on Seattle. His models show a positive impact for young households; Guo and Bhat (2007) also finds a positive estimate for households not owning a car. Additionally, they report a negative influence of the fraction of residential land-use in close proximity for all households types, which supports a preference for mixed land-use. Meanwhile Pinjari *et al.* (2009) reports an opposite effect for the same study area, and observes that homogeneous regions are favoured when using commercial fraction and land-use mix as variables in their residential location choice models. They do not differentiate between household types in their study, which leads us to expect that mixed land-use is valued by households that favour urban areas, young households and households not owning a car, whereas other household types seem to prefer a more homogeneous neighbourhood. A further differentiation with household with children is still outstanding. Within the investigated strand of literature on residential location choice, the usage of the interaction between built environment and choice behavior is still limited as compared to mode choice (Cervero and Kockelman 1997) and car ownership (Chu 2002).

Points of interest

Points of interest (POIs) form the distribution of functions with a relevance for the public, which can be introduced by urban planners and policy-makers or form a reaction of the market; they can be placed in buildings or outside and have an abstract representation as spatial point. Examples are educational facilities as schools or recreational facilities as sport fields. Certain types of POIs as administrative functions or retail are indicators of centre structures as well.

POIs that have been reported in residential location choice literature can be generally grouped into the categories: education, transportation, leisure, retail and service, urban centres. Variables on POIs appear in two forms in our reviewed literature: as distance from a location in form of network distances, absolute distance or travel costs or as the number of POIs within a certain neighbourhood.

Education

A common observation for various study areas is that residents like the proximity of educational facilities. Pinjari *et al.* (2009) mention a positive impact of the density of schools in a zone, although not being of high significance. Axhausen *et al.* (2004) and Vyvere *et al.* (1998) observe a similar effect, as they report a negative sign for distance to schools. None of the reviewed studies mentions the use of other educational facilities or the differentiation of catchmetn areas.

Service and retail

Service functions have been explored by Zondag and Pieters (2005); they find service density in a zone to increase the residential utility for all household types tested. Only non-single household with employees do not show a significant effect; however, this group reacts on accessibility for all travel purposes, which can be assumed to be partly correlated. Lee and Waddell (2010a) use the logarithm of jobs in the neighbourhood as variable in their model and proves this to have a positive effect, while Guo and Bhat (2007) find that lower income households and single households are more likely to reside near employment centres when using employment accessibility. We therefore can expect the service density and retail density have a common positive influence on residential choice.

The influence of the proximity to retail facilities has only been explored within stated preference surveys for the literature we reviewed. All studies find them to have a positive effect on the residential utility. Vyvere *et al.* (1998) show that the distance to grocery shops is valued when less than 500m and the distance to a shopping centre is appreciated when being under 5 km. Kim *et al.* (2005) find the travel costs (which we can expect to be correlated to distance) to increase the probability of moving out of a location and to reduce the utility of a residential location.

Recreation and sport
Pinjari *et al.* (2009, 2011) report on the use of variables representing the density of sport and recreational facilities for residential location choice. They find the number of physically active recreation centres to significantly enhance the utility of a location. The number of natural recreational centres has only a minor influence for households with bicycles. No other study we reviewed has used measurements on the sport and recreational facilities, but based on this report and own observations we expect the proximity to sport and recreation facilities to have a positive effect as long as they are not noisy.

Transportation facilities
de Palma *et al.* (2005) report that the proximity to subways in Paris is valued, while the proximity to railway stations is disliked by households. They state that this is related to the noisy environment of railways stations and the multiple retail services that group around subway stations. Vyvere *et al.* (1998) report a positive effect of proximity to bus-stations in their stated preference survey in Louvain-la-Neuve. Habib and Miller (2009) report on individual car traffic and find the proximity to highway exits to be of positive effect. We thus expect that proximity to local public transport stops (bus, tram, subway) is of relevance for residents, while transport facilities for long distance transport, such as railway stations and highway exits, might only be valued by certain groups of persons and otherwise are avoided due to their noisy environment.

Urban characteristics & centre
Several studies include explicit variables describing urban characteristics in the residential location choice. Andrew and Meen (2006) find a relation between life cycles in their study on London and show that households tend to move toward the city core when they are young, and later move away from the city. de Palma *et al.* (2005, 2007) show a significant negative value for a variable representing the centre of Paris. Kim *et al.* (2005) includes a variable for city settlement in their stated preference survey of Oxfordshire. Their residential model shows a clear tendency to move out of the city as well. Belart (2011) and Bürgle (2006) define the Central Business District in Zurich as a spatial reference point and report a trend of all households to move away from this one spatial point. Axhausen *et al.* (2004) include the distance to Mittelzentrum and Oberzentrum . Their model show the distance to a mayor centre (Oberzentrum) to be valued by residing households. Meanwhile the proximity to smaller centre structures (Mittelzentrum) is reported to have a positive impact on the residential utility by Axhausen *et al.* (2004), leading to the expectation that dense urban areas are generally disliked while local centre structures are valued within residential location choices. A second study with a more differentiated approach has been performed by Zondag and Pieters (2005). They distinguish four types of urban characteristics (urban centres, urban neighbourhoods, local village centres and local village green neighbourhoods) and explore the impact on the residential location choice for different household types. Although several of their models report a significant effect of urban characteristics, it is not possible to distinguish general tendencies, which might rely on the classifications of households. The other discussed studies lead us to the assumption that urban characteristics have an effect on residential location, closely related to life cycles of households. Households tend to move away from the city core during this later phase. However all these models have the problem of missing reproducibility of the variables representing urban characteristics. It is not clearly defined what it is the extent of the city core, or where the single spatial reference point is placed. The definition of a point or zone as urban core is less convincing as they are not reproducible nor defined. Instead models should investigate to capture these characteristics by other spatial variables, e.g. built density, density of services and public transport density. Also, a further differentiation of household types should be envisioned as young households tend to favour urban areas.

Socio-economic environment

Variables describing the socio-economic environment are considered in most studies and are commonly available through census data and statistics of municipalities on an aggregated level, often being administrative boundaries. Attributes used to characterize a location are aggregated household statistics (size, age, income, origin, children, workers) and employment rates. This group of variables thus represents the non-fixed configuration of the urban landscape, which is constantly in flux. Land-use simulation models as UrbanSim have their focus on simulating these dynamics and use data models that can capture these measurements. In their analyse of neighbourhood perception Guo and Bhat (2007) find that social-economic and demographic composition values have smaller spatial extent of influence than land-use variables

Many of the socio-economic attributes are defined as interaction terms with characteristics of the households and thus could also be seen as household attributes. For the classification in this review we consider them to be of spatial relevan

Population density
An attribute which is used in most residential choice models is population density. de Palma *et al.* (2005, 2007) implement three different formulations in their model specification: absolute density, log of density and change of population density. The results are the same in the two reports, showing the absolute population density to be dominating and being negative. Also other studies find the population density to have a negative effect for residential choice (Kim *et al.* 2005; Lee and Waddell 2010a; Weisbrod *et al.* 1980). Zondag and Pieters (2005) differentiate between household types in the Netherlands; household generally dislike population density, while employed, one-person households feel attracted by densely population areas. Guo and Bhat (2007) find a negative impact on residential utility for small families and high income households but a positive impact for all other households, including young households and single-person households and thus supports the previous observations. Bürgle (2006) shows that especially young households are attracted by areas of with high populations density in Zurich. As mentioned previously this type of households has a general preference for urban areas, mixed land-use and a high density of dwelling units. Pinjari *et al.* (2009, 2011) finds a general positive influence using a logarithmic formulation of population density, but a negative impact for senior households, and households with children.in the untransformed formulation. Zolfaghari *et al.* (2012) reports a positive influence of population density for all households. Concluding, it can be expected that population density is generally disliked by households, especially by families but that certain types of households, namely young households and single person households value the population density due to their preference for urban areas.

Household types
Several authors have used the share or density of same household types as the relocating household as attributes in their location choice models. de Palma *et al.* (2005, 2007) shows that most households in Paris tend to search the proximity to households of same size; only two-person households do not prefer this, albeit with low significance. Guo and Bhat (2007) find a positive influence. In addition, they use the difference between household size and average zonal household size as variable, which has a negative impact, underlining the previous results. This latter approach also has been mentioned in other studies Pinjari *et al.* (2009, 2011); Weisbrod *et al.* (1980); Zolfaghari *et al.* (2012), all finding a negative impact when household sizes differ. Waddell (2006) introduces two variables: households of similar age and households of similar size. He is the only author who reports a significant effect for elder families (age higher than 40). It is expected that a negative preference for dense populations is leading to this counter-intuitive observation, as he does not including population density in his model nor the share of same households (but the total number of households).

Lee and Waddell (2010a,b) demonstrate that families like the proximity to other households with children using the percentage of families as an explaining variable in their location choice models. Bürgle (2006) observes a similar effect when including density of children in the neighbourhood as variable.

Two studies further report that young households tend to cluster for the study areas of Seattle and Paris (de Palma *et al.* 2005, 2007; Lee and Waddell 2010a). The latter study additionally shows a negative influence of a high density of young households, which might represent correlation effects to the preference of young households for urban areas. Pinjari *et al.* (2009, 2011) and Weisbrod *et al.* (1980) use the fraction of senior persons as explaining variables on residential choice of senior households and find a positive influence as well. Waddell (2006) uses the density of households with same age instead of the fraction. His models only shows a significant positive influence for young single-person households. Similar to his observations on household size, we expect that the integration of a density instead of a percentage, captures effects of population density that have a negative influence.

de Palma *et al.* (2005, 2007) classify households according to the number of workers and assume that households segregate based on this characteristic. In a first study in 2005 they find a significant positive impact, an effect which is not observed in the later study again.

These reports lead us to assume that household prefer to locate around households of same household type concerning age, size and the presence of children. It thus is an essential variable for modelling residential location choice and needs to be integrated as an interaction term. The share of households with a same type tends to be valued positive and the difference of the average size to the own size can be expected to be negative.

A question that has not been explored within the reviewed literature and demands further research is the following: to what point socio-economic variables are causal explanatory variables at all; is it that similar households tend to group or that they have the similar housing, spatial and locational preferences and tend to segregate.

Household origin and race
Various studies have observed segregation effects when defining ethnic groups or origin of households as a variable in residential location choice models. de Palma *et al.* (2005, 2007) differentiate between households with foreign head and French headed households in the Paris region and demonstrate that households with a foreign head tend to group while households with a French head perceive the vicinity of foreign households negative. Waddell (2006) finds the very same effect for minority household and white households. While minorities favour the neighbourhood of minorities, white households will avoid the neighbourhood of minorities. Pinjari *et al.* (2009, 2011) and Guo and Bhat (2007) differentiate between various ethnic groups within their models and observe that all households tend to locate near households of the same group. We expect that households tend to group according to their ethnic group and origin and that this variable is of importance in residential location models.

Household income
Several studies used income groups as explaining variable in their residential location choice models and proved this to have a significant effect. Weisbrod *et al.* (1980) observes a preference of high income households to locate around other high income households, while de Palma *et al.* (2005, 2007) observe a grouping of low income households in his models. Also Zondag and Pieters (2005) show that middle or high income households like to reside in the neighbourhood of the same income group, but find a preference for low income households to reside near middle income households. Their models also differentiates between household types and show that retired seniors prefer middle income neighbourhoods and dislike high income neighbourhoods in the Netherlands. Waddell (2006) includes interaction terms for low, middle and high income distribution according to household size and age. His models show a positive influence for all income groups, but being of significance only for some of the low and high income household types. Further studies Guo and Bhat (2007); Pinjari *et al.* (2009, 2011) include the difference between individual household income to the average zonal income and find to have a highly significant negative value. All these reports show that households have a tendency to relocate around household groups with a similar income.

Housing costs
Clearly closely related to the sales prices or rental costs of the individual unit described in the previous section are the spatially aggregated housing costs. Studies observe that these housing costs have a significant and negative influence on the utility of a location (Axhausen *et al.* 2004; Pinjari *et al.* 2009, 2011; Zondag and Pieters 2005). Srour *et al.* (2002) includes the average lot value in a zone (land price) as explaining variable and finds it to be of negative impact on the residential utility of a location. Guo and Bhat (2007) include the ratio of income to the average housing price in a zone within their model instead of the absolute formulation, which has a negative sign. As these studies use a zone as choice alternative instead of an individual dwelling unit, this is not surprising. It is not expected that beyond this effect, the average price of a zone will have any explanatory power unless it is used in relation to the individual costs of a residential unit.

Employment
Different approaches have been reported implementing the distribution of employment into residential location choice models. Two studies integrate the unemployment rate in their models and show a negative effect on the utility of a residential location for London and Toronto (Andrew and Meen 2006; Habib and Miller 2009). Other studies use the density of jobs as en explaining variable. Zondag and Pieters (2005) demonstrate this to have a significant positive effect but only for non-single, employed, households of middle or high income, while Srour *et al.* (2002) observe a general attraction for locations with a high job density in Dallas. Pinjari *et al.* (2009, 2011) test this variable for different household types, but do not find any significant effects. Also de Palma *et al.* (2005, 2007) do not find any significant and consistent preferences, as this variable has opposite signs in their two studies. This leads to the expectation that the density of jobs does not have any consistent

effect on residential location choice, but that households avoid locations with a high unemployment rate.

School quality

Kim *et al.* (2005) and Zhou and Kockelman (2008) find school quality to have a positive influence; as it is a non-interaction variable this holds for all households. Chen *et al.* (2008) include school quality as interaction variable for households with children and without children and find it to be slightly stronger for households with children, but having a positive and significant sign for both. However the attribute "school quality" is not clearly specified. Andrew and Meen (2006) include the GCSE-levels and A-level scores (proportion of children obtaining 5+ GCSEs grade; proportion of children obtaining 3+A/AS levels) in their model and find a positive impact of the first variable but also report of a negative impact for the latter. They conclude that households rather send their children to other schools as reaction to poor local neighbourhood schools instead of moving, but also expect that another reason might be the correlation to deprivation index. Weisbrod *et al.* (1980) include the ratio of teachers to students as a quality indicator. This has a positive influence on the residential utility of a location for households with children.

We would assume that school quality is of relevant for households with children or have family plans, even if some studies find this variable to be valued positive be all households. Further research will have to explore the expected correlation with other socio-demographic variables and how (and if) school quality can be measured in a land-use simulation.

Other variables

Some additional variables are only used in a few studies so that a general expectation on their influence and their use within residential choice models is not possible. Weisbrod *et al.* (1980) includes the property tax rate per household as a variable in the model specification, which does have a negative impact with a low significance. Bürgle (2006) also finds a negative impact for the study area of the Greater Zurich Region, though with a high significance, which reflects the local divergence of tax rates and the high competition between municipalities in Switzerland.

Two studies mention the use of rental vacancy within residential choice models. While Zondag and Pieters (2005) report of the positive influence of vacant housing in the neighbourhood for the Netherlands on all household types tested, Bürgle (2006) finds a negative impact using the municipal rental vacancy. A common observation of Andrew and Meen (2006) and Weisbrod *et al.* (1980) is that the crime rate has a significant negative impact on residential location choice. In the reviewed studies, Guo and Bhat (2007) are the sole authors to include the share of owner-occupied housing for residential location choice models and find this variable to be of positive effect for the location utility of owners.

Access & accessibility

Access

The access to work, or commute time. is often included when the workplace of the decision-maker is known and the commuting time can be computed for all alternatives. Generally it is obtained from a transport model and needs to be calculated for every alternative of the choice set. The commute time to work from a location tends to be of negative influence to the residential utility a location (Axhausen *et al.* 2004; Guo and Bhat 2007; Habib and Miller 2009; Lee and Waddell 2010a; Zhou and Kockelman 2008; Zolfaghari *et al.* 2012). If commuting time is not available studies use the commuting distance as network based distance or Euclidean distance instead (Bürgle 2006; Belart 2011; Chen *et al.* 2008; Srour *et al.* 2002), which always has been reported to have the same negative sign, but has a lower significance. Few studies differentiate between commuting time by car, public transport and eventually commuting costs (Kim *et al.* 2005; Pinjari *et al.* 2009, 2011) which all are reported to have a negative impact on the residential utility when being of significance. Commute times by transit are found to be more important than commute times by private transport (de Palma *et al.* 2007). Belart (2011) and Pinjari *et al.* (2009) include the average commuting time respectively commuting distance of all workers in a household further enhances model performance. A weighted approach based on percentage of employees has not prooved to enhance the model though (Belart 2011). To capture the reduced effect of long distance variations, Bürgle (2006) uses a formulation within an exponential function which is also implemented by Belart (2011).

Accessibility

Accessibility is a measurement of the spatial distribution of activities about a point, adjusted for the ability and the desire of people or firms to overcome this spatial separation or 'the potential of opportunities of interaction' (Hansen 1959). Accessibility-calculations are commonly applied to represent local and regional differences within the urban landscape. Five main types of accessibility measures have emerged in literature (Bhat *et al.* 2002): spatial separation, cumulative opportunities, gravity measures, utility measures and time / space measures.

Some studies rely on an aggregated zonal accessibility measure (e.g Guo and Bhat 2007; Zolfaghari *et al.* 2012). A negative impact on the residential utility for San Francisco and London respectively can be observed. The latter study also proves that single households are more likely to reside near employment centres, a preference which already was reported within the population density variables and which can reflect preferences for urban areas. Srour *et al.* (2002) measures accessibility by means of different methods. They find that the cumulative opportunity accessibility index provides best results for residential location choice, based on statistical significance and behavioural interpretation and find a preference for accessibility to employment. Belart (2011) differentiates between accessibility based on the public transport network and the road network in Zurich. Jobs accessible through public transport network prove to have a significantl and positive effect for the location choice of households having no car, meanwhile the accessibility based on road network shows a negative influence on residential utility for car owners. Bürgle (2006) also observes a preference households not owning a car to live in places of high accessibility to the population in the same region. Zondag and Pieters (2005) calculate a logsum measure for all travel purposes and find accessibility to be a significant variable in the move-stay choice. Accessibility of a specific location is only of significance for employed non-single households and retired single households, being disliked by the earlier and favoured by the latter. Chen *et al.* (2008) include accessibility to open space and find that this is positive for households with kids. However, the role of accessibility is smaller than that of other factors such as income and other household related factors. Ben-Akiva and Bowman (1998) consider a logsum measure based on the activity schedule of an individual and find this measure not to improve their model, similar to Eliasson (2010), who consider the direct utility from the optimal activity pattern. Besides accessibility to employment, he considers accessibility to services and shops and concludes that this is necessary to include. Lee and Waddell (2010a) apply a time space prism measure to calculate shop accessibility and a highly disaggregated work travel time measure. These two measures are highly significant and have a relatively large influence. They argue that these factors are being considered by decision-makers and should be included in residential location models. Accessibility to recreational facilities is found to be disliked by residing households (Srour *et al.* 2002; Zolfaghari *et al.* 2012) Zondag and Pieters (2005) state that the effect of accessibility is marginal in comparison to the attributes of the location and the residential unit, as local differences in accessibility are small.

The differentiation into socio-demographic groups (car owners, single person households) seems to capture observations that have already been done before: some types of households tend to have a preference for urban areas. Eventually the differentiation of these can lead to common observations, but the reviewed literature does not allow on a clear statement.

3.4 Summary

Throughout the literature review a common behaviour for several attributes of different study areas can be found. A final overview on the variables found to be relevant within the reviewed literature is given in figure 2.

On the level of the household, commuting distance variable is reported to be of significant negative impact in most models, as well as the distance to the previous location, which is rarely used, but highly recommended. The latter variable both serves as a proxy for distance to social contacts but also to known amenities

The characteristics of the residential unit include price and size. Price, in the form of rental costs or sales price, has negative influence in all studies, with different taste preferences between owners, renters and shared accommodation. Floor space is valued by all households; single person households tend to prefer less and bigger rooms in comparison to other household types. Different housing typologies, such as single family (detached) and multifamily houses, is mainly of relevance for families with children.

Built environment variables have not been included in most studies reviewed. Especially density measures and land use mix seem to be a promising way to capture the extent of urban areas. Land-use mix has been implemented by predefining the extent. A different mix is preferred by households in different life cycles. A differentiation between households by age is preferred. Den-

HOUSEHOLD ATTRIBUTES

group	variable	function	radius	reported impact	expected correlation
commuting time	commuting time/distance/costs by car * car available	distance		-	
	commuting time/distance/costs by pt * no car available			-	
individual distances	distance to previous location	distance		-	*social network*
	distance to social network				*previous location*

RESIDENTIAL UNIT

group	variable	function	radius	reported impact	expected correlation
size	floorspace per person * non.single hh	value		+	*rooms per person;*
	floorspace per person * single hh			+	*rooms per person*
	rooms per person * non single hh	value		+	*room density to avoid correlation with floorspace!*
	room per person * single hh	value		-	*room density to avoid correlation with floorspace!*
costs	(price/income) * hh is owner	value		-	
	(price/income) * hh is renter	value		-	
house type	dettached single family * hh with children	value		+	
	multifamily housing * single person hh	value		+	
age	building age	value		-	
	historic building				
	new building	boolean			
quality	*building quality*				

BUILT ENVIRONMENT

group	variable	function	radius[m]	reported impact	expected correlation
built density	log (density of dwelling units in 600m)	density	500	+	*density of population*
	log(density of dwelling units) * hh has children	density	500	-	*density of population*
	log(density of dwelling units) * (young hh or single hh)	density	500	+	*density of population*
open space	share of open space or unbuilt space	spatial ratio	500	+	
	share of water	spatial ratio	500	+	
land use	share of commercial land use * (young hh or single hh or hh without car)	spatial ratio	500	+	
	share of commercial land use * hh has children	spatial ratio	500	-	
	share of industrial land	spatial ratio	500	-	
	share residential land	*spatial ratio*			
urban character	distance to urban center (CBD) * (nonsingle hh or hh has children)	distance		+	*accessibilty of jobs*
	distance to urban center (CBD) * (young hh or single hh)	distance		-	
	distance to local center	distance		-	
network/noise	buffer to arterials and railways (noise)	boolean		-	
	structural denisty(amount of network links per km²)		1000		

character of neighborhood	*number of historic buildings*
	number of new buildings OR share (newbldg/totblg)
sunshine	*sun exposure of topographie*
topographie	*view of mountains or lakes*

POINTS OF INTEREST

group	variable	function	radius	reported impact	expected correlation
education	distance to school	distance		-	
	distance to university	distance			
	distance to kindergarten.	distance			
service and retail	density of retail	density	500	+	*share of commercial landuse*
	density of service	density	500	+	*share of commercial landuse*
	distance to grocery	*distance*			
	distance to shoping center	*distance*			
sport and recreation	density of sport activity centers	density	500	+	
	density of natural recreation centers	density	500	+	
transport	distance to local transport	distance		-	
	distance to station * no car owner	distance		-	
	Very close proximity to station (noisy envrionment)	value		-	
	distance to highway exit * car owner	distance		-	
	density of local transport	*density*			

SOCIO-ECONOMIC ENVIRONMENT

group	variable	function	radius	reported impact	expected correlation
population density	population density	density	1000	-	*built density*
	populations density * (young hh or single hh)	density	1000	+	*built density*
	share of vacant apartments				
household type	share of hh with same size	spatial ratio	1000	+	
	share of hh with same age	spatial ratio	1000	+	
	share of hh with children * hh has children	spatial ratio	1000	+	
	share of hh with same income cat	spatial ratio	1000	+	
	share of hh of same ethnic group/origin	spatial ratio	1000	+	
employment	unemployment rate	value		-	
crime	crime rate	value		-	

ACCESS AND ACCESSIBILITIES

group	variable	function	radius	reported impact	expected correlation
sociodemographic	accessibilty (cumulative opportunities) of jobs * (no car or young or single)	acc		+	*distance to centre*
	accessibiltys (cumulative opportunities) of jobs * car	acc		-	*distance to centre*
POIs	accessibility (cumulative opportunities) of shops	acc		+	*density of retail*

Variables written in italic way: alternative that have not been reported in the reviewd literature, but is expected to be of influcence due to comparable attributes.

Figure 2: Summary of reported model results and proposal of generic model setup with expected influences

sity of the road and rail network is an indicator for noise and can be implemented as a buffer. This variable shows a negative impact for all household types. Guo and Bhat (2007) state that the spatial perception of built space in terms of land use is wider than for variables describing the socio-economic environment. This is an aspect that demands further research.

Points of interest form a group of variable that has appeared in several studies, showing consistent behavioral preferences in different regions. Education, service, retail and local transport facilities are being valued by all household types, i.e. distance is disliked and density of these facilities is appreciated. Preferences for stations and highway entries and exits vary, depending on mobility tools of the household and the noise level these transportation facility. Distance to the Central Business District (CBD) and to urban settlements show similar preferences: young households tend to favour the proximity while all others prefer to locate away from the urban core.

Variables describing the socio-economic environment are included in the majority of the studies. The population density seem to reflects the location of urban centres and eventually might be correlated to the urban centre. Population density is often preferred by single households and young persons. All household types tend to dislike the unemployment rate in a location and the crime rate. Studies in which the share of household types is considered find a significant effect of households types to segregate based on size, age, children, income and the origin or ethnic group.

Accessibility to employment does not show a consistent taste preferences, despite being addressed in several studies often with a high level of significance. We expect that these different preferences reflect different preferences for mobility and would suggest to distinguish accessibility by with transport mode as done by Belart (2011). It is expected that these accessibility correlates with variables describing urban structure sucg as dwelling density or population density, reflecting network design guidelines. This would explain the sometimes reported preference of single households for highly accessible locations. Accessibility to shops has shown to be valued by all household types, but is partly captured by the density of shops.

4 A data model for land-use simulation

In the previous section ideas were expressed for variables to include in land-use simulation, based on variables included in a wide range of residential location choice studies. Earlier land-use simulation approaches use the rather aggregated statistical spatial level of grids and zones. The availability of spatial data in a disaggregated level not only allows to include different data source for residential location choice models, but also demands a disaggregation of the data models for usage in land-use simulation.

UrbanSim's data model reflects this development. Figure

The classification of this review implies the differentiation of variables into (spatial) objects and agents. While objects represent entities of the built environment that are influenced by decision-makers, such as urban planners and policy-makers, agents represent individuals acting within the urban environment and making use of the objects. This usage of objects reflects individuals preferences as modeled with choice models and resulting in the distribution of agents over space. From a planners perspective this implementation of spatial objects strongly increases the benefits the usage of land-use simulation for decision support, as it allows for the inclusion of changes in the urban configuration and test their effect on the distribution of agents.

Figure 4 gives an overview of a proposed data model as basis for land-use simulation. In comparison the other data models presented, we aim to add enterprises and add points of interest as well as networks. This will be elaborated upon in the ensuing.

Based on various results of different studies we propose to make a distinction between buildings, networks and point of interest. Furthermore, land entities (parcels) are needed to represent open space. This makes it possible not only to include characteristics of the unbuilt space, but to integrate land-ownership and regulations on future construction eventually by means building location choice model, e.g. height regulations, density regulations and other rules for the spatial configuration.

While building represent the physical location of jobs and households, the disaggregation of buildings into a discrete set of residential units is recommended to allow for the modeling of characteristics of individual residential units (Schirmer *et al.* 2013). Whether office space

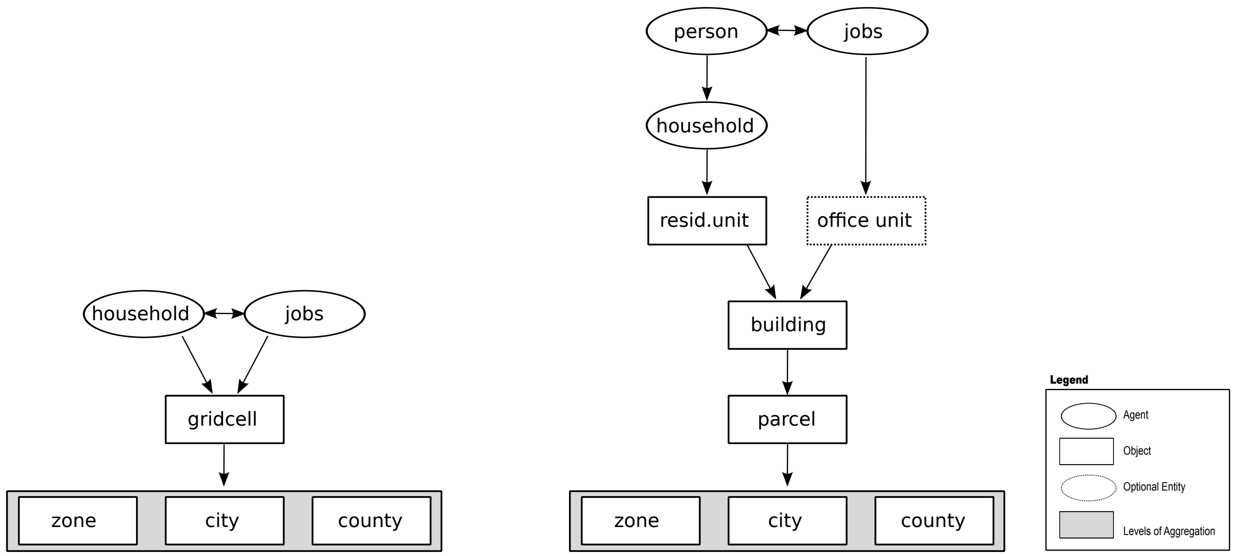

Figure 3: Data models of UrbanSim in the gridcell version (left) and the implemenation of the parcel version as implemented in the Zurich Case Study of the project "SustainCity" (right side)
fig:UrbanSim

needs to be specified as office units will need to be addressed in further research considering employment location choice.

A distinction has been made between several classes of points of interests in this review: education, transportation, service, retail, recreation and sport. A further subdivision of these into categories representing the scale of influence is expected to be useful for the modeling purpose, e.g. a university has a larger catchment area than a primary school or a kindergarten.

Networks form the linkage within the urban landscape and are needed for accurate distance and travel time calculations by different modes, e.g. the distance to workplace. As Guo and Bhat (2007) show, modeling these distances and the extent of a neighborhood is best done by network-based distance calculations. In addition these networks also represent a source of noise and emissions, which have an impact on residential location choice.

In addition to these spatial objects we differentiate between agents describing households and enterprises, which are an aggregation persons and jobs. Modeling these as single agents allows to model trade-offs within the household or enterprise and the use of individual attributes, e.g. the distance to workplace .

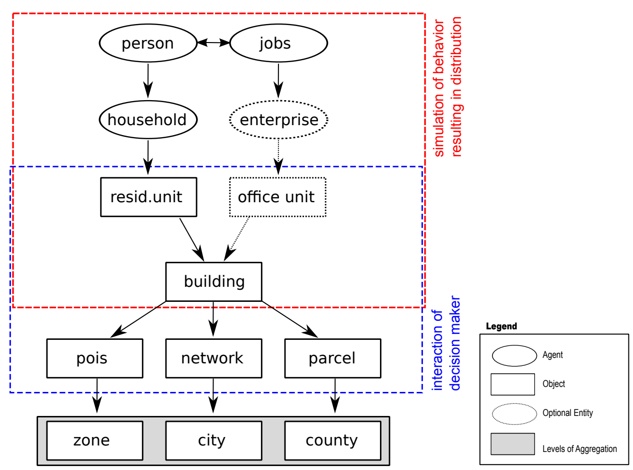

Figure 4: Recommended data model based on reviewed location choice models

5 Outlook

In order to enhance the usability of disaggregated land use and transport simulations for planning and decision support systems, future work will benefit by using a common data model and generalized model specifications, at least as a starting point for the setup of a land-use simulation. To these means, this paper presented an exhaustive review of residential location choice models. Despite differences in methodology, model specification and formulation of variables a consistent preference structure for various variables can be found. However, based on these differences, we acknowledge that this review cannot be seen as proof of concept and further research is required. A first step to do so has been reported in a completive paper on the study area of Zurich (Schirmer *et al.* 2013)

In the context of the SustainCity project (www.sustaincity.org) we aim to investigate which model specification and variables can be used in an initial set-up of a transport land-use simulation. This will be done for the Zurich area, as a wide range of disaggregated data sources are available in the context of the project. As a second step, we aim to implement the data model as discussed in this paper, making a distinction between points of interest, networks and parcels. The ongoing work will show how useful this datamodel ist and what differentiations it will need.

However, by adding these further spatial typologies to a transport land-use simulation a new challenge arises. Several studies have pointed out the relevance of different types of points of interest. Whereas some of the points of interest should be introduced externally by the analyst, such as public transport stops and roads, other points of interest are the result of the interaction between consumers, retailers and developers and will need to be simulated. The challenge is to further capture these processes and subsequently capture the urban processes we observe in everyday life.

References

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

Anas, A. 1982. *Residential Location Markets and Urban Transportation*. New York: Academic Press.

Andrew, M. and G. Meen. 2006. Population structure and location choice: A study of london and south east england. *Papers in Regional Science*, 85(3):401–419.

Axhausen, K., D. Scott, A. König, and C. Jürgens. 2004. Locations, commitments and activity spaces. In M. Schreckenberg and R. Selten, eds., *Human Behaviour and Traffic Networks*, pp. 205–230. Berlin: Springer.

Bürgle, M. 2006. Residential location choice model of the Greater Zurich area. In STRC, ed., *6th Swiss Transport Research Conference*. Ascona.

Belart, B. 2011. *Wohnstandortwahl im Grossraum Zürich*. Master's thesis, IVT, ETH Zurich, Zurich.

Ben-Akiva, M. E. and J. L. Bowman. 1998. Integration of an activity-based model system and a residential location model. *Urban Studies*, 35(7):1231–1253.

Ben-Akiva, M. E. and S. R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge: MIT Press.

Bhat, C. R., S. L. Handy, K. Kockelman, H. S. Mahmassani, A. Gopal, I. M. Srour, and L. Weston. 2002. Development of an Urban Accessibility Index: Formulations, Aggregation, and Application. Working Paper Report 7-4938-4, University of Texas Austin, Austin.

Cascetta, E., F. Pagliara, and K. W. Axhausen. 2007. Dominance attributes for alternatives' perception in choice set formation. Working Paper 371, IVT, ETH Zurich, Zurich.

Cervero, R. and K. Kockelman. 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3):199–219.

Chen, J., C. Chen, and H. J. P. Timmermans. 2008. Accessibility trade-offs in household residential location decisions. *Transportation Research Record*, 2077:71–79.

Chu, Y.-L. 2002. Automobile ownership analysis using ordered probit models. *Transportation Research Record*, 1805:60–67.

de Palma, A., K. Motamedi, N. Picard, and P. Waddell. 2005. A model of residential location choice with endogenous housing prices and traffic for the paris region. *European Transport\Trasporti Europei*, 31:67–82.

de Palma, A., N. Picard, and P. A. Waddell. 2007. Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region. *Journal of Urban Economics*, 62(2):204–230.

Eliasson, J. 2010. The Influence of Accessibility on Residential Location. In F. Pagliara, J. Preston, and D. Simmonds, eds., *Residential Location Choice: Models and Applications*, Advances in Spatial Science, pp. 137–164. Berlin: Springer.

Eluru, N., I. N. Sener, C. R. Bhat, R. M. Pendyala, and K. W. Axhausen. 2009. Understanding residential mobility: A joint model of the reason for residential relocation and stay duration. *Transportation Research*

Record, 2133.

Gordan, I. 1992. Modelling approaches to migration and the labour market. In A. Champion and A. Fielding, eds., *Migration processes and patterns 1: Research Progress and Prospects.* London: Belhaven.

Guevara, C. A. 2005. *Endogeneity and Sampling of Alternatives in Spatial Choice Models*. Ph.D. thesis, Massachusetts Institute of Technology, Boston.

Guo, J. Y. and C. R. Bhat. 2002. Residential location choice modeling: Accommodating sociodemographic, school quality and accessibility effects. Technical report, Department of Civil Engineering, University of Texas, Austin, Texas.

Guo, J. Y. and C. R. Bhat. 2007. Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography*, 15(1):31–45.

Habib, K. M. N. and E. J. Miller. 2009. Reference-dependent residential location choice model within a relocation context. *Transportation Research Record*, 2133:92–99.

Hansen, W. 1959. How Accessibility Shapes Land Use. *Journal of the American Institute of Planners*, 25(2):73–76.

Iacono, M., D. Levinson, and A. El-Geneidy. 2008. Models of Transportation and Land Use Change: A Guide to the Territory. *Journal of Planning Literature*, 22(4):323–340.

Kim, J., F. Pagliara, and J. Preston. 2005. The intention to move and residential location choice behaviour. *Urban Studies*, 42(9):1621–1636.

Krizek, K. J. and P. A. Waddell. 2003. Analysis of lifestyles choices: neighborhood type, travel patterns, and activity participation. *Transportation Research Part B: Methodological*, 1807:119–128.

Löchl, M. and K. W. Axhausen. 2010. Modelling hedonic residential rents for land use and transport simulation while considering spatial effects. *Journal of Transport and Land Use*, 3(2):39–63.

Lee, B. H. Y. and P. A. Waddell. 2010a. Reexamining the influence of work and nonwork accessibility on residential location choices with a microanalytic framework. *Environment and Planning A*, 42(4):913–930.

Lee, B. H. Y. and P. A. Waddell. 2010b. Residential mobility and location choice: A nested logit model with sampling of alternatives. *Transportation*, 37(4):587–601.

Lowry, I. S. 1964. A model of metropolis. Technical report, RAND Europe.

Müller, H.-P. 1992. *Sozialstruktur und Lebensstile-Der neue theoretische Diskurs über soziale Ungleichheit*. Frankfurt am Main: Suhrkamp.

McFadden, D. 1974. Conditional logit analysis of qualitative choice-behaviour. In P. Zarembka, ed., *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.

McFadden, D. 1978. Modeling the choice of residential location. In A. Karlqvist, ed., *Spatial Interaction Theory and Residential Location*, pp. 75–96. Amsterdam: North-Holland.

Otte, G. 2005. Entwicklung und test einer integrativen typologie der lebensf□hrung f□r die bundesrepublik deutschland. *Zeitschrift für Soziologie*, 34(6):442. ISSN 0340-1804.

Pinjari, A. R., C. R. Bhat, and D. A. Hensher. 2009. Residential self-selection effects in an activity time-use behavior model. *Transportation Research Part B: Methodological*, 43(7):729–748.

Pinjari, A. R., R. M. Pendyala, C. R. Bhat, and P. A. Waddell. 2011. Modeling the choice continuum: an integrated model of residential location, auto ownership, bicycle ownership, and commute tour mode choice decisions. *Transportation*, 38(6):933–958.

Schirmer, P., M. Van Eggermond, and K. Axhausen. 2013. Measuring location in residential location choice - an empirical study on the canton of zurich. Utrecht.

Srour, I. M., K. Kockelman, and T. P. Dunn. 2002. Accessibility Indices: Connection to Residential Land Prices and Location Choices. *Transportation Research Record*, 1805:25–34.

Swait, J. D. 2001. Choice set generation within the generalized extreme value family of discrete choice models. *Transportation Research Part B: Methodological*, 35(7):643–666.

von Thunen, J. H. 1826. *Der isolierte Staat in beziehung auf Landwirtschaft und Nationalokonomie, Gustav Fisher, Stuttgart. English edition: The isolated state (trans: Wartenburg CM (1966), edited by Hall P).* Oxford: Pergamon.

Vyvere, Y., H. Oppewal, and H. Timmermans. 1998. The validity of hierarchical information integration choice experiments to model residential preference and choice. *Geographical Analysis*, 30(3):254–272.

Waddell, P. 2006. Reconciling household residential location choices and neighborhood dynamics. *Working*

Paper.

Walker, J. L. and J. Li. 2007. Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9(1):77–101.

Wegener, M. 2004. Overview of Land Use Transport Models. In D. A. Hensher, K. J. Button, K. Haynes, and P. R. Stopher, eds., *Handbook of Transport Geography and Spatial System*, pp. 127–146. Oxford: Elsevier.

Weisbrod, G., M. E. Ben-Akiva, and S. R. Lerman. 1980. Tradeoffs in residential location decisions: transportation versus other factors. *Transport Policy and Decision Making*, 1(1).

Zöllig, C., R. Hilber, and K. W. Axhausen. 2011. Konzeptstudie Flächennutzungsmodellierung.

Zhou, B. and K. Kockelman. 2008. Microsimulation of residential land development and household location choices: bidding for land in austin, texas. *Transportation Research Record*, 2077:106–112.

Zolfaghari, A., A. Sivakumar, and J. W. Polak. 2012. Choice Set Pruning in Residential Location Choice Modelling: A Comparison of Sampling and Choice Set Generation Approaches in Greater London. *Transportation Planning and Technology*, 35(1):87–106.

Zondag, B. and M. Pieters. 2005. Influence of accessibility on residental location choice. *Transportation Research Record*, 1902:63–70.