ETH zürich

ETH Library

Integration of a capacity-constrained workplace choice model

Recent developments and applications for an agent-based simulation in Singapore

Journal Article

Author(s):
Vitins, Basil J.; Erath, Alexander; Axhausen, Kay W.

Publication date:
2016

Permanent link:
https://doi.org/10.3929/ethz-b-000104463

Rights / license:
In Copyright - Non-Commercial Use Permitted

Originally published in:
Transportation Research Record 2564, https://doi.org/10.3141/2564-01

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

Integration of a Capacity Constrained Workplace Choice Model: Recent Developments and Applications for an Agent-Based Simulation in Singapore

Date of submission: 2015-07-19

Basil J. Vitins
Future Cities Laboratory, #06-01 CREATE Tower, Singapore 138602
phone: +41 44 633 27 02
fax: +41 44 633 10 57
basil.vitins@ivt.baug.ethz.ch

Alexander Erath
Future Cities Laboratory, #06-01 CREATE Tower, Singapore 138602
phone: +65 8409 8023
fax: +41 44 633 10 57
alexander.erath@ivt.baug.ethz.ch

Kay W. Axhausen
Institute for Transport Planning and Systems (IVT), ETH Zurich, CH-8093 Zurich
phone: +41 44 633 39 43
fax: +41 44 633 10 57
axhausen@ivt.baug.ethz.ch

Words: 6139 words + 3 figures + 1 table = 7139 word equivalents

ABSTRACT

Destination choice models can be embedded in travel demand models to understand travel behavior and forecast future scenarios. Utility-based destination choice models can account for individual travel behavior and are therefore suitable for agent-based models. Accounting for destination capacities is also in line with the agent-based theory. Therefore, this paper addresses the possibility and impact of introducing capacity constraints and their effect on choice behavior, as well as the feasibility to apply such an approach in agent-based micro-simulations with individual specific characteristics for each agent.

This paper describes a workplace choice model and its application in a large-scale simulation of Singapore. One methodological and one technical achievement are highlighted in this paper. Regarding the methodological progress, capacity limitations are introduced at workplaces, and a methodology based on shadow prices is employed to consider the capacity limitations within the choice model applications. A robust optimization methodology efficiently assigns the commuters to workplaces while respecting the individual preferences of commuters. The technical achievement profits from recent computational advances. The workplace choice model is estimated with a comprehensive utility function on a large data set and 10^3 destinations and applied and cross-checked on the entire population. A reasonable model fit and robust parameters are achieved by obviating sampling techniques. The proposed methodology is efficiently applied for commuting trips with minor travel time changes compared to a reference sample.

INTRODUCTION

Destination choice is a central and challenging problem in travel choice modeling. It can be characterized by the discrete choice of a specific person between different destination alternatives under a given situation. Compared to other discrete choice situations, such as mode choice, destination choice has to deal with very large number of choice alternatives specifically located in space. Destination choice models can be based on utility-maximization models and are therefore especially suitable for agent-based modeling as they can capture the heterogeneity among individual behavior. E.g. employment type or income can be used as independent variables when assigning individuals to workplaces of a specific economic sector. Among various destination choice models, this paper focusses on workplace choice as part of the overall travel demand modeling process for a transport model of Singapore. Commuting trips are of major importance due to their large share during the peak hours. Planning long-term investment in transportation infrastructure and spatial development will profit from workplace choice modeling and forecasting in general.

Within demand modeling, also capacity constraints at choice alternatives are essential whenever the market supply is limited for agents in a choice situation. Capacity constraints should therefore be considered in the (utility-maximization) workplace choice model. Capacity constraints are fully integrated in the doubly-constrained gravity-based trip distribution. Capacity constraints are only partially considered in utility-maximization models, and additional adjustment is required due to capacity constraints under certain circumstances and depending on the specific model. This paper therefore discusses issues regarding capacity constraints at destinations and proposes an efficient methodology to specify constraints in utility-based models and their applications. However, the paper omits a comparison between gravity- and utility-based approaches as it was done in e.g. *(1)*.

Recent advances in computational performance allow parameter estimations for choice models on large (full) sample sets with large numbers of destination alternatives and complex utility functions. Additionally, they allow model applications on populations as large as $\sim 1.9 \cdot 10^6$ commuters, as given in Singapore.

The aim of this paper is twofold. The workplace choice model should account for capacity constrained destinations. Additionally, recent computational and technical advances should be exploited to achieve robust parameters and reasonable model fit, and applications in a large-scale transportation simulation. Therefore, the main research questions query the impact of introducing capacity constraints and the effect on choice behavior; as well as the feasibility to apply such an approach in a large-scale application while accounting for heterogeneity in destination choice.

Destination Choice Modeling

Much research about transport related destination choice has already been conducted in the past. This section highlights selected past achievements on (1.) the methodological side and (2.) the data and sampling side. Regarding (1.), many papers have explored different discrete choice methodologies. E.g. *(2)* applied an MNL model with a differentiated utility function. An MNL model can be estimated computationally efficiently for destination choice purposes and therefore can allow a large number of alternatives and complex utility functions. *(3)* provided a seminal paper on residential choice and several model formulations. Other researches applied complex choice methodologies and destination choice models to especially account for unobserved similarities between the destination alternatives. *(4)* summarized and compared different GEV

(generalized extreme values) models in practise with relaxed IID assumptions compared to the MNL model to capture at least some of the unobserved similarities between the alternatives, namely the Nested Logit (NL), Cross-Nested Logit (CNL), Generalized Nested Logit (GNL), the spatial correlated model (SCL) and a combined model. All these models were developed in the past and have specific pros and cons regarding model characteristics, complexity and computational burdens. Regarding destination choice, the SCL *(5)* deserves special attention due to the consideration of adjacent zone pairs which might be spatially correlated. In the meanwhile, the Generalized Spatially Correlated Logit model (GSCL) can be added which considers the spatial distances between all the alternatives *(6)* compared to adjacent alternatives in the SCL. Beside the model theory, also the solution algorithms are crucial when it comes to model applications. They strongly depend on the model formulations. A closed-form model formulation can be solved with direct maximum likelihood techniques, compared to other formulations which require more challenging numerical and simulation based approaches *(e.g. 7)*.

Regarding (2.), sampling techniques in destination choice models have been widely discussed due to the large choice sets. *(3)* provided techniques for sampling. Also *(8)* and *(9)* proposed sampling methods among others, and found that a rather large number of observations are needed to achieve reasonable model parameter values. *(9)* suggested to draw 1/8 of the full choice set size as a minimum and 1/4 as a desirable sample share in the case of their MNL models, and state that non-MNL models are even more demanding regarding the required sample size. Recently, there is a trend to avoid sampling techniques of the alternatives in destination choice models and to directly calibrate model parameter with the entire set of alternatives. This has the advantages that no sampling techniques is required and even complex models can be applied without worrying about the sampling strategy. This is done e.g. in school choice models where pupils are assigned to different schools *(10)* where the choice set is the set of potential schools in a given area. When it comes to workplaces, normally the set of potential workplaces is considerably larger. The workplaces are aggregated into zones to achieve reasonable calculation times, as calculation time approximately linearly increases with the number of alternatives *(e.g. 11)*.

Capacity Constraints and Shadow Prices

Assuming a utility-based workplace choice model accounting for generalized travel costs, quality of the destination, personal and situational variables, it can be argued that workplace choice reflects the empirical behavior including a certain market competition for e.g. more centrally located workplaces. However, the competition and market equilibrium for attractive workplaces and low rents take place on a firm level with a different utility function. Therefore, we cannot a priori assume a market clearing situation in the case of workplace location choice, as it might be in case of housing and residential location *(e.g. 3, 12)*. Certain issues might lead to over- and under-saturated workplace locations after applying a utility-based destination choice model:

1. Generalized travel costs are essential in destination choice models. They are negatively weighted within the utility function, meaning that nearer locations are preferred compared to destinations further away. Assuming a situation with spatially separated residential and workplaces, it might be that the travel cost weights are lower when the calculations are based on an RP survey than in reality. If commuters would really be able to compete for nearby and attractive workplaces, the weights for travel costs might be different.
2. Destination choice models obviously heavily depend on survey data. Whenever the survey data is stratified or biased with regards to spatial attributes, it can be assumed that certain

(commuting) behaviors are not well captured for a given stratum within the model. It is possible to overcome this problem by estimating parameters for the under-represented strata, but a certain parameter bias can be expected to persist after parameter estimation, and therefore a bias in the destination choice model applications.

3. (a) Recent very elaborate discrete choice models can deal with complex choice situations, e.g. spatial correlation (see also Section "Destination Choice Modeling"). However, it is still possible that not all the correlations are actually captured within the model. Therefore, systematic errors might still be possible.
   (b) The perception of generalized travel costs might not be proportional to the actual costs. E.g. *(2)* included a differentiated distance function, composed of six terms. One can assume that the perception can be approximated with complex functions, but residual errors might still remain and bias the modeling outcome.
4. Even if a perfectly fitted model is assumed based on a representative data set, parameter values might change in scenario applications. In case of scenarios with changing destination capacities, commuters might not completely be reassigned to all the changing destinations due to the influence of the generalized travel costs, which is an alternative specific variable but completely ignores capacity. An example is a spatially restricted area like the island of Singapore.
5. Assuming an imperfect model due to one of the reasons outlined above, even a small over-saturation can lead to errors when evaluating and interpreting results, e.g. for business developments of the over-saturated and under-saturated destinations, and the related traffic flows.

The idea of shadow prices is proposed in the following to account for many of the above issues in destination model applications. Shadow prices are applied differently in various studies. In economics, shadow prices are used to estimate unknown costs of a certain good or alternative. In the following, shadow prices are applied as an additional impedance for attractive but limited alternatives. Shadow prices should therefore reflect the constraints of a certain alternative. In microeconomics and productivity optimization, the constraints would be the price $p \geq 0$, a stock of X with sold units x where $0 \leq x \leq X$, and the objective $\max(px)$. Customers buying the units x optimize their utility u which is as well subject to time and budget constraints. We therefore have (at least) two dependent optimization problems with various constraints. Any resource is considered as constraint (e.g. time, units) if the amount that customers would like to use exceeds the amount available for them. In the case of an inefficient market, it might be possible that the demand exceed the supply due to distributional effects, and a shadow price needs to be implemented to still be able to solve the optimization problem *(13)*. This idea of the shadow prices is transferred and adjusted in the following for the destination choice situation.

In destination choice, shadow prices can be assigned to designated destinations and can be regarded as additional impedance for persons choosing these destinations. Shadow prices therefore are able to account for constraints of certain alternatives which cannot be captured with model parameters or the error term (ϵ) distributions of the choice model. Current literature provides little discussion of shadow prices and their effect in choice models. Shadow prices are applied in certain studies, but the detailed methodology or reference is missing as well as the definition and effect of the prices. *(14)* used shadow prices for park and ride lots to identify the best parking lot, as well as *(15)*. *(16)* applied shadow prices in a mode choice model for CO_2 pricing. *(12)* modeled the housing market and provided a detailed definition of the shadow prices. *(12)* stated that the the prices in the housing marked might clear the housing market depending on the specific situation. They presented algorithms for constant and variable demand in the

housing market. The case of workplace models described in this paper is different (compared to e.g. *(12)*) and especially office rents are not fully mirrored in the utilities of the employees. Therefore, the employee's utility is not directly affected when choosing a certain workplace compared to an identical workplace with lower rents.

Some researcher were also focussing on the joint travel and residential costs and theoretically formulated mathematical programs. E.g. *(17)* theoretically introduced shadow prices similar to the proposed approach for welfare maximization of home location changes under infrastructure changes, including housing rents. *(18)* modified the utility function and transportation and residential costs to improve residential choice and the usage and impact of transportation.

Compared to the advantages of applying shadow prices in destination choice models, there are also disadvantages:

1. Shadow prices cannot replace and omit variables which are ignored in the model, but which might be related for the model quality. The model will not be policy sensitive in future scenario for these variables.
2. Shadow prices might be replaced by alternative specific constants in a destination choice model. However, these constants are not scenario and policy sensitive for model applications.

It can be summarized that the issues listed at the beginning mainly remain open based on our literature review, and a more detailed definition and knowledge of the effect of shadow prices is relevant for future applications.

DATA AVAILABILITY FOR THE CASE STUDY OF SINGAPORE

The destination choice model was estimated based on an RP household travel survey conducted for the entire island of Singapore. Singapore is located in Southeast Asia and covers an island of 43 km length and 23 km width, with a permanent resident population of $3.8 \cdot 10^6$ and total population of about $5.4 \cdot 10^6$ in 2014. Singapore is one of the wealthiest countries in Asia regarding GDP/capita, and the number of cars is increasing continuously (35.1% of all households have at least one privately registered car available and 48% have a vehicle available *(19)*).

Household Interview Travel Survey

The Household Interview Travel Survey *(19)* is conducted every four to five years and the last survey was conducted between June 2012 and May 2013 by the Land Transport Authority (LTA) of Singapore (further referred to as "HITS 2012"), including Singaporean citizens, permanent residents and legal immigrants residing in Singapore. Data of 9'635 households, 35'714 persons, 70'984 trips and 85'880 stages where collected in HITS 2012, providing a substantial sample of the population and their travel behavior. 12'292 trips to workplaces are reported in HITS 2012 excluding weekend trips, as well as trips to schools by students. 20.5% of all commuters drove to work by car, 5.9% as car passenger, 64.2% with public transportation; the remaining share includes company busses (5.6%), motorcycle (3.9%), taxi (3.6%) and other modes. In HITS 2012, origins and destinations are known by their zip codes. In the case of Singapore, nearly every building has its own zip code (generally a single high rise building), leading to approximately 163'000 zip code entries including spatial coordinates for the entire country. This has the advantage that the coordinates are known for every trip origin and destination.

For the estimated choice model, it is important to know that the model is generated for use

with a synthetic population (see Section "Methodology") with a limited number of descriptive variables. Therefore, certain variables are ignored during model estimation due to the fact that they are not available for the synthetic population. Occupation types and individual income are considered as essential variables for choice modeling. Both person-related variables are available in the synthetic population and can be implemented in generalized utility function, however, individual income was favored due to potential future income changes and adaption in future scenarios. Income is subdivided into 12 categories from 0$ to above 8'000$ for every person.

Spatial Data

The main purpose of the spatial data within the destination choice model is to provide information about the destination quality, such as number of workplaces. Additionally, spatial attributes can be linked with personal attributes in generalized utility functions.

The Singaporean Master Plan (MP) *(20)* contains detailed information on the current and planned land use on the island and includes the medium term spatial plans with a 10-15 years planning horizon, compared to the Concept Plan with a longer horizon and the feasibility studies with shorter time horizons. The MP divides the entire island into ~11'000 zones comparable to a parcel size, and assigns different land use types. Furthermore, it defines transport infrastructure such as roads and stations. The various land use types reported in the MP are aggregated to reduce the number of model parameters. 1'169 zones are used for spatial aggregation of the workplace into a reasonably large choice set, however, not all the zones contain workplaces (e.g. forests, water bodies). The following list includes the different land use types:

- Business
- Commercial
- Residential
- Transportation
- Services
- Recreation
- School
- Open

The number of workplaces in Singapore is not captured by any census, and therefore is derived separately through reported travel patterns and the designated land use types of the destinations. The workplaces are aggregated on a zonal level for this study. The detailed description of the workplace distribution in space is provided in *(21)*. Travel time serves as a major variable for generalized travel cost estimation. GPS-calibrated car travel times and transit travel times are available from the current detailed but aggregated model of the Land Transport Authority. They could be replaced with updated travel times from the entire simulation itself whenever it is completed.

METHODOLOGY

The workplace choice model is embedded in a demand generation process with the purpose to provide daily plans for an agent-based micro-simulation for the MATSim platform *(22)*. MATSim simulates the physical movements within the given infrastructure network, as well as optimizes travel utility related to departure time, secondary location, travel mode and joint trips. Prior to the simulation, additional models are applied to generate the entire synthetic population,

as well as the work, education, shop, leisure location choice of each agent within the synthetic population.

The synthetic population is based on a Bayesian network approach. The Bayesian network approach allows to represent and reproduce the structure of the population. It can be described as an interference of a multivariate probability distribution, where a set of variables characterizes the demographic and socioeconomic information of individuals and households, based on sampled observations and aggregated marginal distribution from the census. *(23)* described the synthetic population and underlying methodology in detail. The synthetic population generated for Singapore contains the variables dwelling type, ethnicity, car availability on a household level and age, gender, income employment, car license and citizenship on a person level, all categorical variables. All the categories match the categories of HITS 2012. In addition, all households are assigned to a specific zip code, which means that the are well-defined in space and therefore can serve as a starting point for commuters.

In addition to the data generated in the synthetic population, the choice of owning a car as a household, and having a license as a person is additionally modelled by discrete choice models to predict changes based on future potential scenarios. The decision of going to work is done by applying an MNL model as well, which is based on the available data of the synthetic population (35.2% of the population go to work on weekdays based on the census). The choice of going for work and location choice is modeled sequentially.

The workplace choice model is described in detail in the remaining subsections below. The descriptions of secondary location choice and additional trips for leisure and shopping are determined after the workplace location choice and excluded in this paper. Regarding trip chaining, it is important to add that the reported trip chains in HITS 2012 might contain less sub-trips compared to other countries. The share of home-work-home (h-w-h) trips is approximately 43%, home-education-home (h-e-h) 27%, and home-leisure-home (h-l-h) 9%. We assume underreported very short trips in HITS 2012, however, this issue is out of scope for this paper. Based on the reported data, is is considered to apply a straightforward work destination model and apply secondary location choice consecutively after establishing the workplaces choice.

Workplace Choice Model Methodology

Various methods have been discussed in literature, a selection is listed in Section "Destination Choice Modeling". The section also outlines the current trade-off between complexity and reasonable low calculation time. An MNL model approach is chosen for the current case study mainly due to the reasonable calculation time also in case of a large data set, and the reasonable calculation time requirements to implement and estimate complex generalized utility function, with the hopefully small risk of biased parameter values for the explanatory variables due to unobserved spatial correlation (see *(4)* for an estimate on the potential parameter difference between different models). A rather large data set of 12’292 commuting trips is available *(19)*, therefore calculation time is rather high, also given the number of alternatives. We excluded a sampling method and estimated all parameters with the entire data set and all destination alternatives, because the additional overhead in calculation time was manageable and therefore we could avoid sampling uncertainties. If a sampling method would have been chosen, a minimum sample size of at least 300 destinations would lead to somewhat reasonable parameter values. Since calculation time approximately linearly scales with the number of alternatives *(11)*, the additional overhead of 4 times at most is considered as reasonable. Additionally the

sampling itself is costly regarding calculation time, as well as its dependency on the choice model.

The authors are aware of alternative models like a NL with integrated mode choice. However, calculation time increases considerably for a model with ~1169 destination zones. It was considered as more straightforward to focus on the utility function compared to focussing on the mode choice interaction and correlation, also due to the arguments above.

Utility Function

The elaborated utility function is similar to the function proposed in *(2)*. The initially proposed utility function consists of a personalized mode choice log sum term and generalized utility to combine personal attributes (e.g. income) with the given land use types. The distance decay function in *(2)* is replaced with a mode-specific generalized cost decay function. An overview of the utility function for the deterministic utility from origin i to destination j is given in (1), whereas the individual parts of (1) are explained below. The utility function excludes an accessibility term similar to *(2)*. Accessibility mainly describes the distance-weighted quality of the surrounded area and not the destination alternative itself.

$$U_{i,j} = \underbrace{f_0\left(\gamma_{0,m}, c_m, c_m^2, \log(c_m)\right)}_{\text{1. Generalized travel cost decay}} + \gamma_1 \cdot \underbrace{ln\left(\sum_m e^{U_{i,j,m}}\right)}_{\text{2. Mode utility}} + \gamma_2 \cdot \underbrace{f_1\left(Workplaces \times Income\right)}_{\text{3. Generalized utilities}} \tag{1}$$

1. Generalized Travel Cost Decay

The generalized cost (c) decay function consists of a linear combination of mainly non-linear travel cost modifications. Travel time is considered as the main component of generalized travel costs and is therefore implemented in the model. Additionally, a mode-specific (m) estimation is conducted based on car ownership, which is considered as a major influence on travel behavior. Different distance decay functions for different profession types are excluded, as proposed in *(2)*.

2. Mode Choice Log Sum

The mode choice logsum contains the utilities U of each mode m for a given origin-destination relation i-j. The consideration of all modes can be justified by the fact that the workplace is a long-term decision. Even if a person has a car-oriented travel behavior (which is obviously reflected in the logsum), a specific workplace might still be more attractive if it is also accessible via public transport. It has to be added that the mode choice itself is a rather short term decision, because it determines the choice for a single specific trip, and often includes travel time with a linear weight (also in this research). On top, the mode choice term can correlate with the generalized travel cost decay function. Therefore, detailed experiments were conducted with the mode choice utility (but with minor effect on the utility function as described in Section "Results").

A separate mode choice model was estimated to determine the weights of the different variables. Currently, an MNL model is implemented for car (including car passenger), transit, and others. This subdivision follows the purpose of the overall project of developing mainly a

car and public transport model. Minor modes are not considered at this point even though they might be relevant in some model applications.

3. Generalized Utilities for Land Use Types × Personal Variables

The purpose of the generalized utility is to match personal attributes with alternative specific attributes. In case of the land use types, the corresponding workplaces might be linked to a specific person type. Due to the envisaged future scenario applications, income is considered as a more reliable variable to predict compared to the occupation types. Therefore, income is further considered in the destination model. The generalized utilities replace to a certain extent the distance perception separation by income similar to *(1)*.

Shadow Prices for Limited Capacities at Destinations

In the following, shadow prices are defined as dis-utility added to destinations due to capacity restrictions. So the shadow prices are positive, and are negatively perceived by the choice makers. The following three assumptions are made for the calculation of the shadow prices. (1) It is assumed that the number of workplaces at a given destination is known and fixed for the entire time. (2) The demand at all origins is known and fixed, and it does not change because of different saturation levels at the destinations. (3) The model parameters of the choice models are given beforehand. It is clear that assumption (2) is critical in very specific destination choice models, e.g. restaurant choice during evening peak hour. In the case of workplace choice, assumption (2) might be less critical but obviously depends on the economic situation of the case study. It is clear that assumption (3) is also critical, because the model parameters might be influenced by a given saturated market situation. For future research, it would be interesting to know how much the perceived weights / elasticities of the generalized travel costs based on a hypothetic questionnaire differ from the weights / elasticities of the reported trips. Solutions might exist to overcome the additional complexity of (3), e.g. specific questionnaires. For this paper, it is assumed that the model parameters are fixed and based on the RP data available.

The major difference between the well-known gravity model and the utility maximization methodology is the underlying methodologies themselves. The gravity model spreads the demand (which is generated mostly for a predefined spatial unit or zone) over all destination zones (which are more or less attractive for the origin zone and their demand segments). The utility maximization methodology determines a probability for every traveling person / agent and every destination according to the destination quality, personal taste and situation (see e.g. *(24)*). Now, both approaches have to deal with capacity restrictions at destinations. Whereas the gravity model forces the agents to travel to under-saturated destinations, the discrete choice model implements capacity limits to a certain extent in the utility function. However, the utility function consists of additional terms beside capacity-related variables, and the parameter fitting does not exclusively optimize the destination saturation per se.

In the following it is assumed that there is a certain balance between the workplaces provided by companies and authorities, and the work destinations of the working population. On the one side, it is certainly impossible to have more workers at a designated place than actual workplaces. On the other side, it might be possible to have empty workplaces not taken by any employees. Therefore, a balance is assumed with a certain upper restriction determined by the maximum number of workplaces which should not be violated in the model. It can be assumed that this balance cannot be determined deterministically, and therefore an iterative procedure is derived and proposed in the following to approximate this balanced situation (similar to *(25)*

who addressed parking lot capacities).

Starting with the utility maximization approach, the probability $p_{i,j}$ of choosing an alternative destination j based on origin i is based on the well-known standard logistic regression:

$$g_{i,j} = P_i \cdot p_{i,j} = P_i \cdot \frac{e^{u_{i,j}}}{\sum_j e^{u_{i,j}}} \tag{2}$$

whereas $u_{i,j}$ is the deterministic utility of alternative j. P_i is the number of persons commuting from origin i (a building or a zone unit). Obviously, personal attributes and situational attributes are ignored in (2) by omitting corresponding indices. However, personal and situational attributes might be considered as well.

It can be shown that the following convex minimization problem is equivalent to (2) by applying the Kuhn-Tucker conditions, feasible because the function is partially differentiable *(26)*:

$$\text{Min} \sum_{i,j} g_{i,j} \left(ln(g_{i,j}) - 1 + u_{i,j} \right) \tag{3a}$$

subject to:

$$\sum_j g_{i,j} = P_i \tag{3b}$$

Now, the capacity restrictions (3c) are simply added to the constraints of the optimization problem above ((3a),(3b)). Obviously, the number of employees should not exceed the number of workplaces available (3d). However, this constraint is not necessarily required to solve the problem, as shown later.

$$\sum_i g_{i,j} \leq C_j \tag{3c}$$

$$\sum_k C_k \geq \sum_k P_k \tag{3d}$$

The vectors λ_1 and λ_2 are added as Lagrange multipliers. λ_1 has length i, λ_2 has length j:

$$L(g_{i,j}, \lambda_1, \lambda_2) = \sum_{i,j} \left(g_{i,j} \cdot ln(g_{i,j}) - g_{i,j} + u_{i,j} \cdot g_{i,j} \right) + \lambda_1 \left(\sum_j g_{i,j} - P \right) + \lambda_2 \left(\sum_i g_{i,j} - C \right) \tag{4}$$

Now, the optimality conditions of (4) are ($\lambda_2 \geq 0$):

$$g_{i,j} = e^{-\lambda_{1,i} - \lambda_{2,j} - u_{i,j}} \qquad \text{for all} \;\; i, j. \tag{5}$$

and

$$\sum_j e^{-\lambda_{1,i}-\lambda_{2,j}-u_{i,j}} = P_i \qquad \text{for all } i. \tag{6a}$$

$$\sum_i e^{-\lambda_{1,i}-\lambda_{2,j}-u_{i,j}} \leq C_j \qquad \text{for all } j. \tag{6b}$$

λ_2 is responsible that the capacities are not exceeded and is therefore referred to as shadow price. Unlike (3a) – (3c), the dual problem (6a) – (6b) comes without constraints; λ_1 and λ_2 can be determined by solving (6a) and (6b) iteratively.

For efficiency, all the variables in (6a) and (6b) are transformed: $\alpha = e^{-\lambda_1}$, $\beta = e^{-\lambda_2}$, $U_{i,j} = e^{-u_{i,j}}$, whereas $\alpha, \beta, U > 0$ and $\beta < 1$:

$$\alpha_i \cdot \sum_j \beta_j U_{i,j} = P_i \tag{7a}$$

$$\beta_j \cdot \sum_i \alpha_i U_{i,j} \leq C_j \tag{7b}$$

Efficient Algorithm to Determine the Shadow Prices

Algorithm 1 describes an iterative procedure to determine the shadow prices. t is the threshold

---

Algorithm 1 Shadow price calculation.

$n \leftarrow 0$
$\beta_n \leftarrow 1$
while $C - P < t$ **do**

1. Calculate the demand $g_{i,j,n}$ for each pair i, j based on (5) and (7a):

$$g_{i,j,n} \quad \leftarrow \quad \frac{P_i \cdot \beta_{n-1} \cdot U_{i,j}}{\sum_j \beta_{n-1} U_{i,j}} \tag{8}$$

2. Recalculate the β parameters based on (5) and (7b):

$$\beta_{n,j} \quad \leftarrow \quad \min\left(\frac{C_j \cdot \beta_{n-1}}{\sum_i g_{i,j,n}}, 1\right) \tag{9}$$

3. $n \leftarrow n + 1$

end while
Shadow prices (λ_2) $\leftarrow$ -log(β)
Terminate.

---

value and describes how much the capacity should not be exceeded at a given destination. E.g. $t = -2$ means that the capacity can be exceeded by maximum 2. **Algorithm 1** approximates a balanced situation within an adequate number of iterations where all commuters are assigned to a workplace. The shadow prices λ_2 can therefore be seen as an additional (negative) utility for

each person to respect the capacity constraints. It might be possible that a similar situation is reached by randomly assign weights to the locations, however, it is definitely worth to know how to efficiently approximate a balanced situation where all the commuters find a designated working location.

RESULTS

The results are twofold. The first part describes the estimated destination choice model. The second part outlines the outcome of the shadow price methodology.

Destination Choice Model

The results are described below for each element of the utility function listed above (Section "Utility Function").

1. As described in Section "Utility Function", the non-linear mode-specific (m) generalized cost (c) decay function $f_{0,m}$ consists of $f_{0,m} = \gamma_{0,m}\left(\eta_{0,m} \cdot c_m + \eta_{1,m} \cdot c_m^2 + \eta_{2,m} \cdot log(c_m)\right)$. m accounts in this study for car availability, as mentioned above. c includes travel time as a major component of the generalized costs. Experiments with additional components were conducted such as c^3 and $\sqrt{c}$. Not all of the ηs were clearly significant and an over-fitting is avoided by selecting only the robust elements for $f_{0,m}$. However, it is found that the perception of costs is clearly non-linear. The η-values are listed in Table 1 and were kept fixed during the determination of all other parameters. Additional person or household based variables could be incorporated for m instead of car availability.
2. In all the conducted experiments, the mode choice log sum term correlates with both car and transit time (parameter correlation of 0.7 – 0.8) and does not add to the overall model fit and therefore is ignored in the final model. It is proposed that main influential variables in the mode choice model should be incorporated as personal variables as shown above (as m) or in the generalized utility function as explained below.
3. The workplaces are transformed with the logarithm due to the considerably higher ρ^2 for the model fit and the theoretical necessity *(27)*. Additionally the transformation reduces the correlation with travel time in the case of Singapore, and is is in line with *(2)*. The generalized utilities include the following combinations:
   - Number of workplaces related to business activities × income categories
   - Number of workplaces related to commercial activities × income categories
   - Number of workplaces related to residential activities × income categories
   - Number of workplaces related to service activities × income categories

   The categories "school", "recreation", "transport", "open" are not significant in most of the combinations and are therefore removed from the model. The considered parameters for the generalized utilities are significant through many different model estimations. The considered variables and parameters might contribute to a specific stratum, however, their contribution to the overall fitness of the model is rather low. This is because every parameter can contribute only in its very specific combination, e.g. in the case of calculating the utility for a person with high income travelling to a business zone.

Table 1 shows an overview of the model parameters. The calculation of the model took ~3 days with PythonBiogeme *(28)* on an Intel Xeon 3.07 GHz using 30GB RAM (mainly due to the generalized utilities). Various experiments with different parameter combinations showed robust parameter values.

TABLE 1 Workplace choice model parameters and statistics.

Parameter	value	t-test	p-value
Travel time decay$_{\text{Households with car}}$ ($\gamma_{0\ car}$)	1.19	82.33	0.00
$\eta_{2,\text{car}}$ (single parameter)	-1.0	–	–
Travel time decay$_{\text{Households without car}}$ ($\gamma_{0\ no\ car}$)	1.09	75.84	0.00
$\eta_{0,\text{no car}}$	-0.121	-8.91	0.00
$\eta_{2,\text{no car}}$	-0.688	-30.12	0.00

Generalized utilities:	value	t-test	p-value	value	t-test	p-value
*income**	*business workplaces*			*residential workplaces*		
500	0.0999	9.20	0.02	-0.113	-5.04	0.00
1’250	0.154	15.97	0.00	-0.110	-5.13	0.00
1’750	0.155	21.66	0.00	-0.196	-11.92	0.00
2’250	0.157	23.09	0.00	-0.218	-13.71	0.00
2’750	0.158	18.86	0.00	-0.300	-14.93	0.00
3’500	0.154	21.61	0.00	-0.328	-18.99	0.00
4’500	0.174	19.32	0.00	-0.293	-13.16	0.00
5’500	0.149	13.59	0.00	-0.346	-12.97	0.00
7’000	0.166	13.37	0.00	-0.285	-9.44	0.00
above 8’000	0.128	11.99	0.00	-0.313	-12.90	0.00
	commercial workplaces			*service workplaces*		
500	0.140	11.96	0.00	0.104	7.57	0.00
1’250	0.211	19.91	0.00	0.0353	2.73	0.01
1’750	0.208	26.44	0.00	0.102	10.78	0.00
2’250	0.209	28.11	0.00	0.0739	8.17	0.00
2’750	0.196	21.59	0.00	0.0918	8.33	0.00
3’500	0.211	27.87	0.00	0.0750	8.03	0.00
4’500	0.213	21.83	0.00	0.0628	5.18	0.00
5’500	0.219	19.00	0.00	0.0575	3.99	0.00
7’000	0.252	19.32	0.00	0.00478	0.28	0.78
above 8’000	0.280	26.03	0.00	0.0589	4.39	0.00

Model statistics	value
Number of observations	12’292
Number of alternatives (excluding green spaces, water bodies etc.)	1’087
Init. log-likelihood	-85’416
Final log-likelihood	-76’155
Final log-likelihood (iteration 10)**	-72’603
ρ^2	0.108
ρ^2 (iteration 10)**	0.150

**income*: Monthly personal income in Singapore Dollars.
Based on census data and application of **Algorithm 1.

Figures 1(a) and 1(b) show the travel time and cumulative travel time distributions when the model applied on the synthetic population compared to the travel times reported in HITS 2012. Both figures show that the travel distributions of the entire population match the distribution of the survey.

Shadow Prices

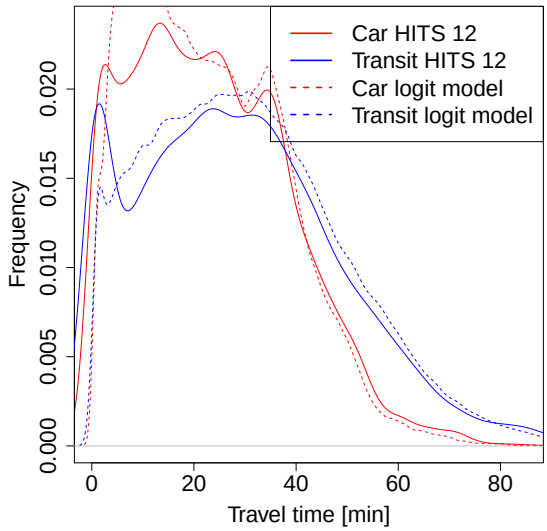
Figure 1 shows the travel time distributions for the synthetic population of Singapore compared to the travel time distribution of the HITS 2012 sample. The destination choice mode is applied as described in Section "Destination Choice Model". The workplaces are independently derived from a separate data source (see Section "Spatial Data"). The algorithm converges to an optimal solution, as shown in Section "Shadow Prices for Limited Capacities at Destinations".

Figure 1 depicts an increasing travel time after 1 and 10 iterations compared to the sampled data, mainly because of three reasons. First, travel time data for very short trips are not captured properly in the data source, and the granularity of the destination zones hampers the determination of precise values for these trips. Second, the distance perception is not captured correctly in the model parameters for very short trips. Third, the workplaces are estimated and therefore a certain additional deviation might take place from real workplace numbers.

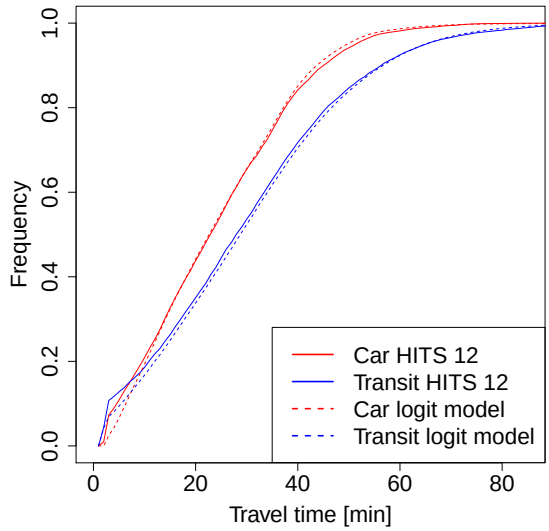
Figure 2 shows the scatter plot for all the destinations and their saturation. Additionally, Figure 2 shows the density of destinations with a given number of workplaces, to reflect the distribution of destinations with different sizes. It is shown that the algorithm converges efficiently to an optimum, and only some outliers are found in iteration 1. This is mainly due to the local distribution of workplaces in Singapore, especially because some bigger industrial zones are located on outlying smaller islands.

Figure 3(a) shows the spatial distribution of the population in Singapore. Figure 3(b) shows the work utilities and Figure 3(c) shows the the shadow prices (λ_2 = -log(β)). It is recognizable that the work utilities and the shadow prices are often complementary. Knowledge about the shadow prices can support location analysis and planning for spatial development. For detailed interpretation, care is required due to the fact that the generalized cost weight (γ_0) is estimated on an RP data source.

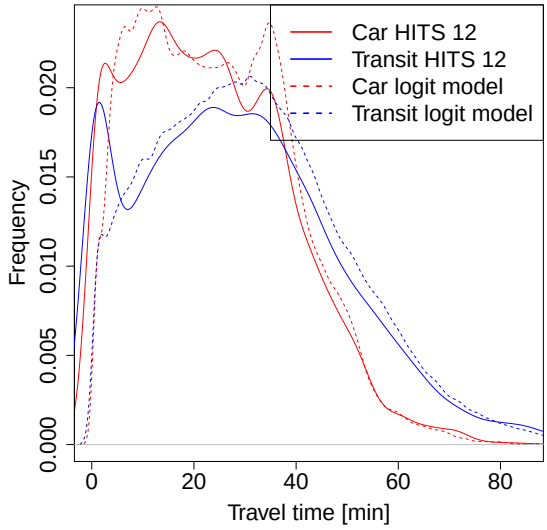
The application of the destination choice model for the entire synthetic population ($\sim 1.9 \cdot 10^6$ commuters) takes about 5.5h / 1.2h. and considerably more memory of about 50 GB (1 thread on an Intel Xeon 3.07 GHz).



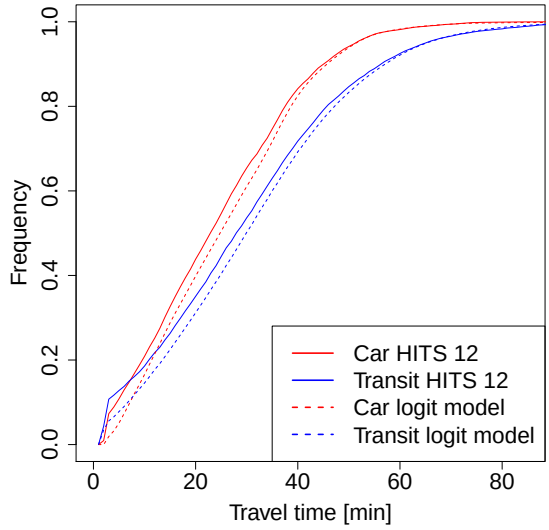
(a) Distribution after applying the choice model (iteration 0 of **Algorithm 1**).



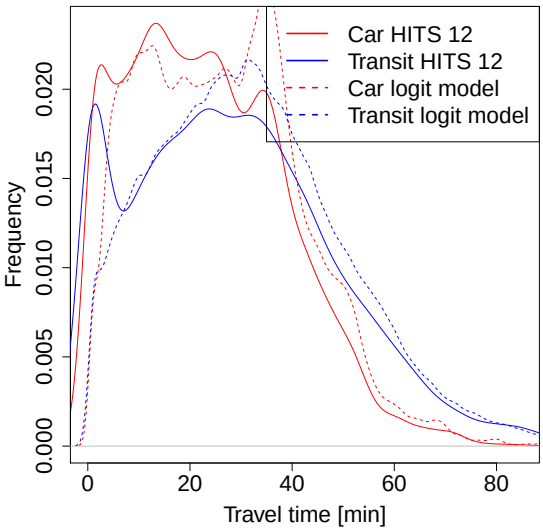
(b) Cumulated distribution after applying the choice model (iteration 0 of **Algorithm 1**).



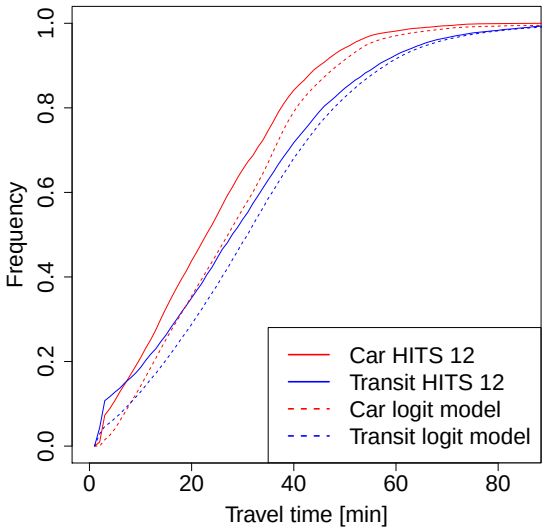
(c) Distribution in iteration 1.



(d) Cumulated distribution in iteration 1.

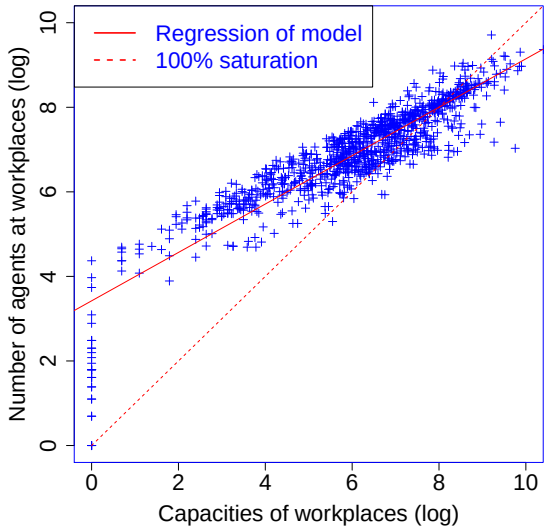


(e) Distribution in iteration 10.

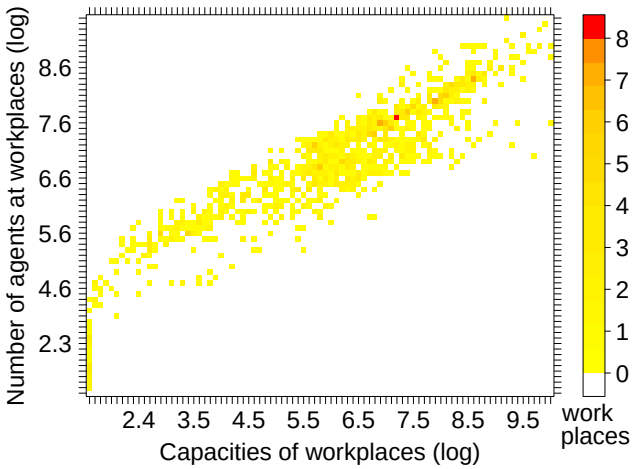


(f) Cumulated distribution in iteration 10.

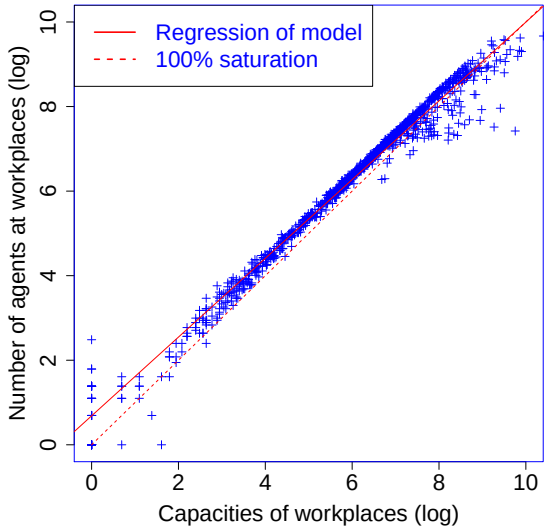
FIGURE 1 Kernel distributions of travel times, and cumulative travel time distributions for the entire synthetic population of Singapore.



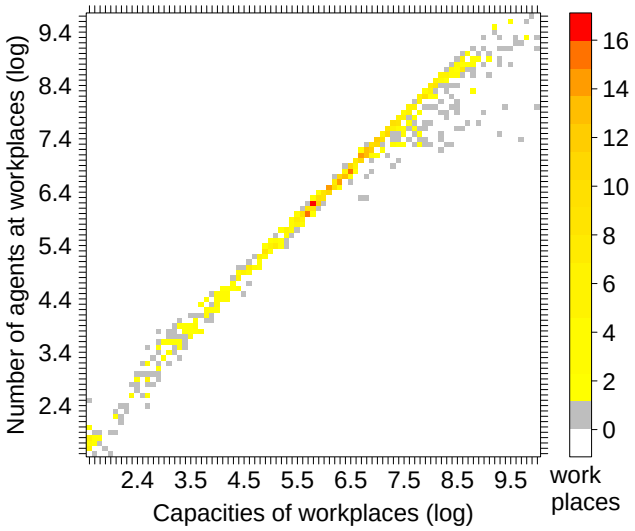
(a) Saturation after applying the choice model (iteration 0 of **Algorithm 1**).



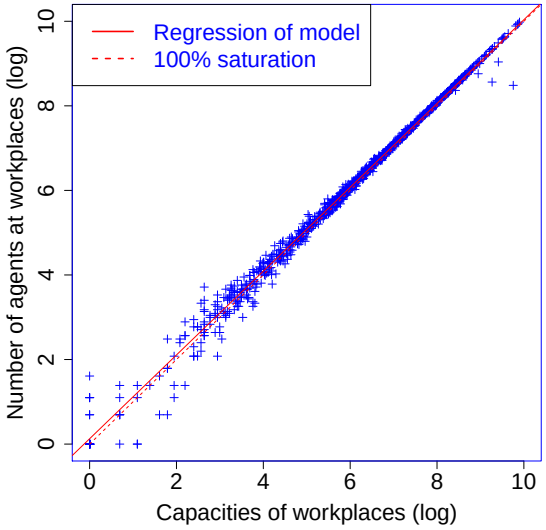
(b) Number of workplaces and saturation after applying the choice model (iteration 0 of **Algorithm 1**).



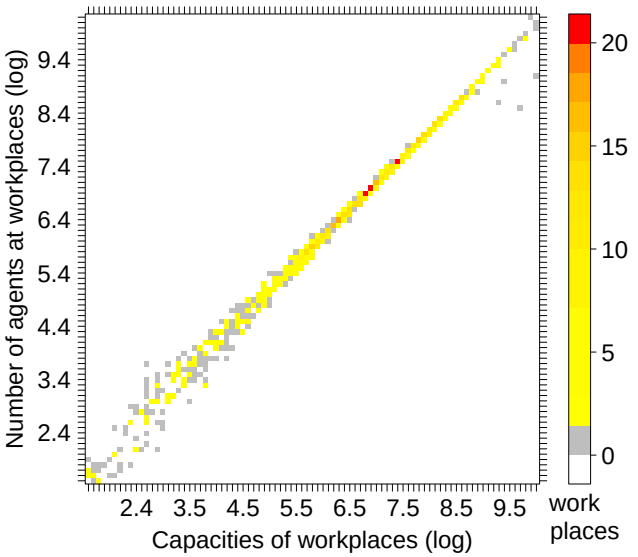
(c) Saturation after iteration 1.



(d) Number of workplaces and saturation iteration 1.

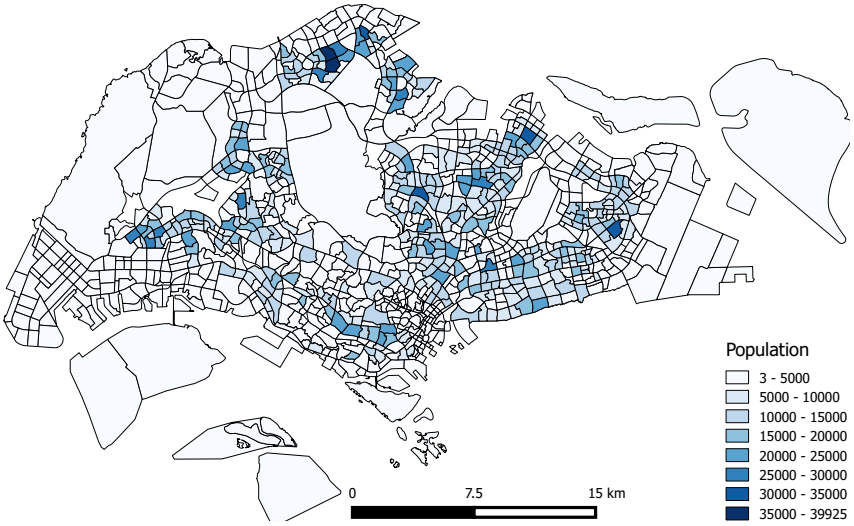


(e) Saturation after iteration 10.

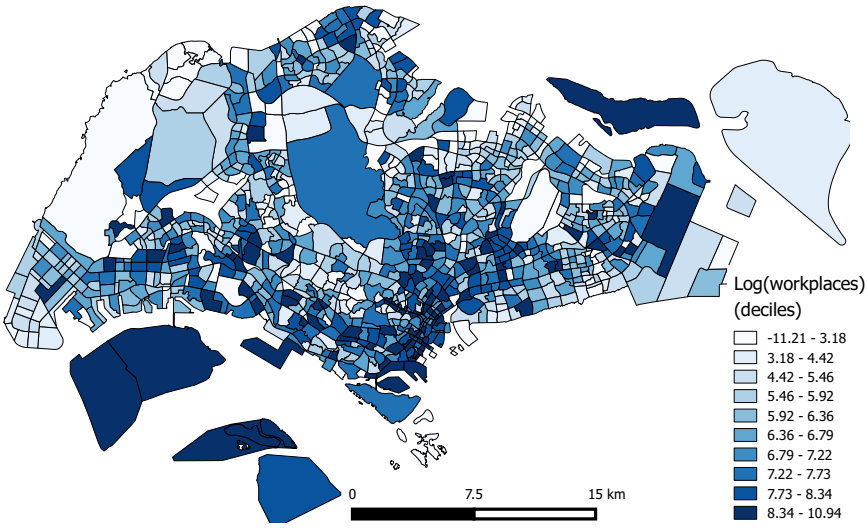


(f) Number of workplaces and saturation iteration 10.

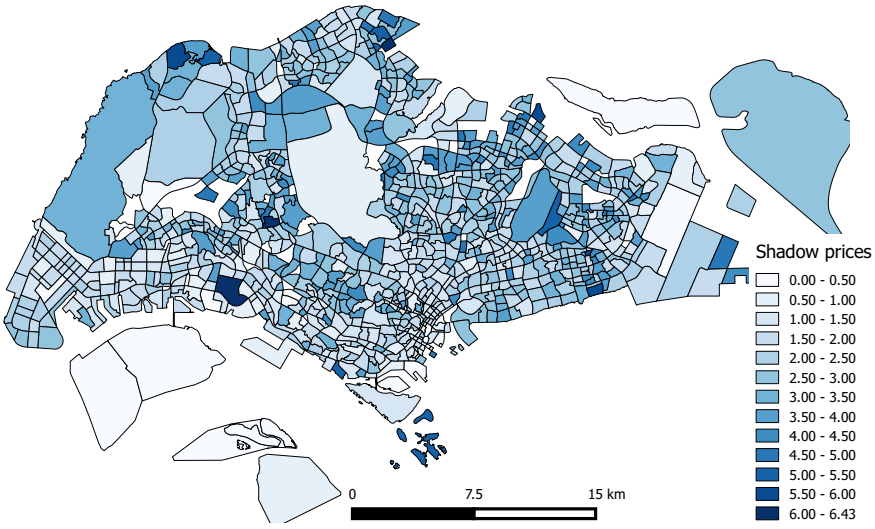
FIGURE 2 Regression and heat map representing the number of workplaces with a certain capacity and saturation for the synthetic population of Singapore.



(a) Population distribution in Singapore.



(b) Workplace distribution.



(c) Shadow prices λ_2 at iteration 10.

FIGURE 3 Population, workplace and shadow prices distributions in Singapore.

SUMMARY AND OUTLOOK

This paper shows that capacity restrictions on destinations can be efficiently implemented in large-scale agent-based models while maintaining choice heterogeneity. It also could be shown that today destination choice models with elaborate utility functions can be applied in larger case study areas based on large data sets and large number of destinations, compared to earlier studies. According to our experiments and results, travel time distribution only slightly change during the optimization of the workplace assignment compared to a reference sample distribution. It is important to know that the proposed shadow price methodology does not replace any destination related variable. The calculation for the shadow prices only holds for MNL models. Applications for other choice models are not proposed so far.

ACKNOWLEDGEMENTS

We thank Sriram Narayanamoorthy for providing advise on the modeling side and Michel Bierlaire for providing Pythonbiogeme version 2.4 *(28)*.

REFERENCES

1. Mishra, S., Y. Wang, X. Zhu, R. Moeckel and S. Mahaparta (2013) Comparison between gravity and destination choice models for trip distribution in maryland, paper presented at the *92nd Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2013.

2. Vovsha, P., S. Gupta, J. Freedman, W. Sun and V. Livshits (2012) Workplace choice model: Comparison of spatial patterns of commuting in four metropolitan regions, paper presented at the *91st Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2012.

3. McFadden, D. (1977) Modelling the choice of residential location, *Working Paper*, **477**, University of California, Berkeley, Yale University, Berkeley, New Haven.

4. Bekhor, S. and J. N. Prashker (2008) GEV-based destination choice models that account for unobserved similarities among alternatives, *Transportation Research Part B: Methodological*, **42** (3) 243–262.

5. Bhat, C. R. and J. Y. Guo (2004) A mixed spatially correlated logit model: Formulation and application to residential choice modeling, *Transportation Research Part B: Methodological*, **38** (2) 147–168.

6. Sener, I. N., R. M. Pendyala and C. R. Bhat (2011) Accommodating spatial correlation across choice alternatives in discrete choice models: An application to modeling residential location choice behavior, *Journal of Transport Geography*, **19** (2) 294–303.

7. Koppelman, F. S. and V. Sethi (2000) Closed form discrete choice models, in D. A. Hensher and K. J. Button (eds.) *Handbook of Transport Modelling*, 235–252, Pergamon, Amsterdam.

8. Frejinger, E., M. Bierlaire and M. E. Ben-Akiva (2009) Sampling of alternatives for route choice modeling, *Transportation Research Part B: Methodological*, **43** (10) 984–994.

9. Nerella, S. and C. R. Bhat (2004) Numerical analysis of effect of sampling of alternatives in discrete choice models, *Transportation Research Record*, **1894**, 11–19.

10. Müller, S., K. Haase and F. Seidel (2012) Exposing unobserved spatial similarity: Evidence from german school choice data, *Geographical Analysis*, **44** (1) 65–86.

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

12. 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.

13. Albouy, D. (2004) Constrained optimization, shadow prices, inefficient markets, and government projects, *Working Paper*, **Economics 131**, University of California, Berkeley, Berkeley.

14. Gupta, S., P. Vovsha, R. Kumar and A. Subhani (2014) Incorporating cycling in ottawa-gatineau travel forecasting model, paper presented at the *5th Conference on Innovations in Travel Modeling*, Baltimore, April 2014.

15. Davidson, B., P. Vovsha and J. Freedman (2011) New advancements in activity-based models, paper presented at the *Australasian Transport Research Forum 2011*, Adelaide, September 2011.

16. Hammadou, H. and C. Papaix (2014) Estimating shadow-prices of urban transport mode choice for a second best pricing of CO_2 emissions, paper presented at the *93rd Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2014.

17. Beckmann, M. J. and J. P. Wallace (1969) Evaluation of user benefits arising from changes in transportation systems, *Transportation Science*, **3** (4) 344–351.

18. Los, M. (1979) Combined residential-location and transportation models, *Environment and Planning A*, **11** (11) 1241–1265.

19. Land Transportation Authority (2013) Household Interview Travel Survey 2012, webpage, `http://www.lta.gov.sg/`.

20. Urban Redevelopment Authority (2015) Master Plan 2008, webpage, `https://www.ura.gov.sg/maps/`.

21. Ordonez Medina, S. A. and A. Erath (2013) Estimating Dynamic Workplace Capacities using Public Transport Smart Card Data and a Household Travel Survey, paper presented at the *92nd Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2013.

22. MATSim (2015) Multi-Agent Transportation Simulation, webpage, `http://www.matsim.org`.

23. Sun, L. and A. Erath (2015) A Bayesian network approach for population synthesis, paper presented at the *4th Symposium of the European Association for Research in Transportation*, Copenhagen, September 2015.

24. Daly, A. J. (1982) Applicability of disaggregate models of behaviour: A question of methodology, *Transportation Research Part A: Policy and Practice*, **16** (5) 363–370.

25. Spiess, H. (1996) A logit parking choice model with explicit capacities, *Research Report*, EMME2 Support Center, INRO, Aegerten.

26. Horst, R. (1979) *Nichtlineare Optimierung*, Carl Hanser Verlag, Munich.

27. Train, K. E. (1986) *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand*, MIT Press, Cambridge.

28. Bierlaire, M. (2003) BIOGEME: A free package for the estimation of discrete choice models, paper presented at the *3rd Swiss Transport Research Conference*, Ascona, March 2003.