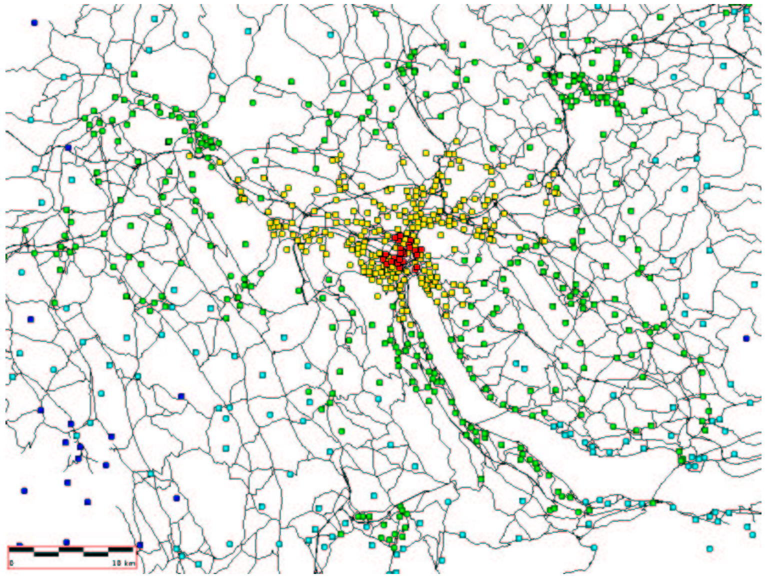


Integration of time of usage in large-scale simulations

Fabrice Marchal, CoLab ETH Zurich

Conference paper

Moving through nets:
The physical and social dimensions of travel
10th International Conference on Travel Behaviour Research
Lucerne, 10-15. August 2003

Integration of time of usage in large-scale simulations

Fabrice Marchal
Computational Laboratory (CoLab)
Swiss Federal Institute of Technology
Zürich

Phone: 00 41 1 632 56 79
eMail: fmarchal@inf.ethz.ch

Abstract

This paper presents an attempt to integrate dynamic traffic models and location choice models in a more consistent way than the four-step planning scheme. Two temporal horizons are considered. On the long run, individuals select the location of their primary activities based on local land use data and on travel costs. Mode choice is omitted. On the short run, the car travelers select the departure time and route for the primary activity. On the short run, the travel demand is disaggregated at the individual level. On the long run, users characteristics are aggregated. A methodology is developed that can be used to bootstrap multi-agents simulations in an efficient way. The emphasis is put on the provision of an operational module. Therefore, the paper focuses on the technical details of the implementation of this methodology and on its applications to large-scale systems. Preliminary results are presented for the area of Zurich, Switzerland.

Keywords

dynamic assignment, location choice, traffic simulation

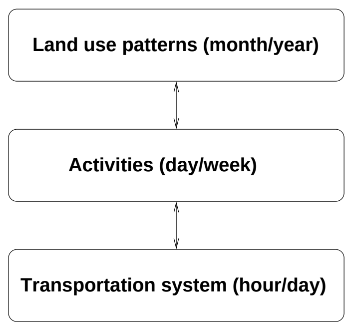
Preferred citation

Marchal F. (2003). Integration of time of usage in large-scale simulations, paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

1. Introduction

Current issues in the study of travel behavior are concerned about the temporal organization of human activities and the journeys that take place between them. Surveys and empirical works are available that unveil part of the real decision processes of users. These are extremely useful to the model designer looking for an improvement or a replacement solution to the classical four-step planning scheme. A major drawback from the four-step scheme is the absence of any time scales: the sequence of trip generation, trip distribution, mode choice and trip assignment is not specified in any time reference. Furthermore, the feedback loop between the upper and lower levels is often neglected or at best implemented using ad hoc rules since the dynamics of the decision update is not obvious. Activity based demand models are a promising replacement solution. These models describe travel demand by modeling how users organize their daily activities and their schedules. Basically, travel demand is considered as being derived from activity demand that is, in turn, spatially dependent on the land use patterns as pictured on Figure 1.

Figure 1: Three layers of models with different time scales



The shift from the trip-based approach to the activity-based approach leads to several problems. First as stated by (Ben-Akiva and Bowman, 1998), *the fundamental problem [...] is combinatorial* in the sense that the dimension of the choice set of activity decisions is extremely huge (10^{16}). The difficulty is the availability of methodologies to model the decision process. Discrete choice models have been very successful in travel demand models but space and time are continuous variables. This can be somewhat relaxed using an adequate discretization. However, the operational requirements of discrete choice models impose a complete enumeration of the choice set, which is intractable in the general case. Obviously some simplifying assumptions are needed to turn Figure 1 into a realistic operational model. A second problem is to provide a consistent way to move data across the different models. Multi-agent simulations like those proposed by (Waddell *et al.*, 2003) and (Raney *et al.*, 2003) address this problem in an elegant way: Individual agents keep their intrinsic characteristics throughout the whole chain of models, without any aggregation performed.

The full implementation of the global model described by Figure 1 will certainly be computationally intensive (as for the TRANSIMS implementation on supercomputers, see (TRANSIMS, 2003)). The purpose of this paper is to provide a simple methodology to avoid starting the whole architecture from scratch data in order to compute the equilibrium for all the modules. Indeed, we can think of starting with a reduced version of the overall model that will be used to initialize the feedback responses between the three different layers. Another goal of this paper is to reduce the data requirements of transportation planning models by avoiding the expensive collection of trip tables. We restrict ourselves to the case of private transportation and remove, for the time being, mode choice from the context.

An important aspect in our view is the interpretation or pre-processing of data in travel behavior research. Typically, the studied area is cut into Transportation Analysis Zones (TAZ) and local land-use attributes are aggregated over these zones. In practice, the administrative boundaries (counties, states, etc.) are often used to define these TAZ. When these divisions are too coarse, for instance in a dense urban area, the practitioner usually divides further the TAZ into smaller pieces. Unfortunately, this pre-processing of data is ambiguous and may differ from one application to another. Therefore, it sounds appealing to use data with the highest resolution available. For travel demand, the highest resolution level is usually that of the local population census. For instance, URBANSIM recognizes that argument and uses 150mx150m cells to describe the land use properties of the system. URBANSIM provides a loose coupling of land use models and transportation models. The specification of the interface with the travel model is that TAZ have to be defined as a group of cells. However, the definition of those groups is left to the user.

In the four-step planning scheme, the trip distribution is often performed using gravity-like models (Ortuzar and Willumsen, 1994). The trip-table T_{ij} is computed using

$$T_{ij} \doteq \frac{E_i A_j}{(tt_{ij})^{\alpha}} \,, \tag{1}$$

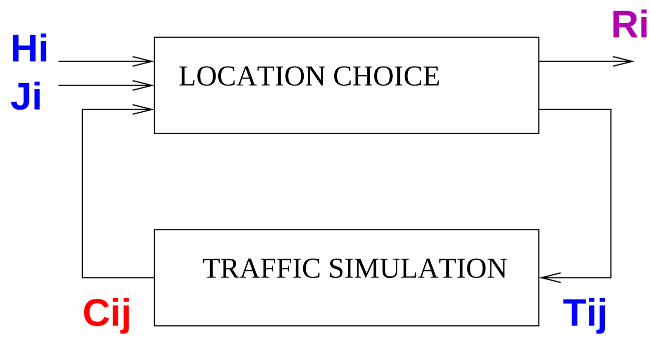
where E_i is the emission of zone i, A_j is the attraction of zone j, tt_{ij} is some impedance related to the journey from i to j (e.g. distance or travel time) and α is the exponent that is adjusted to fit the observations. The main disadvantage of this method is that the α exponent does not have an easy economic interpretation. As a consequence, it requires a real-world measurement of a subset of T_{ij} to fit the α exponent. The dimension of T_{ij} is n^2 where n is the number of TAZ. The collection of a fraction of n^2 estimations is often prohibitive and unreliable. Moreover, the lifetime of these data might be rather short in practical cases because of the evolution of the travel demand over time. In practice, once trip tables T_{ij} have been estimated, they are progressively updated using data with a lower dimensionality such as traffic counts.

The alternative solution proposed here is that of Figure 2. The trip distribution is performed by a location choice model that uses typical raw census data: the number of active residents H_i and the number of jobs J_i in each cell. The dimensionality of the data is n once they are aggregated over the TAZ. Another input is the transportation cost C_{ij} that is unknown at the beginning of the simulation. The location choice model produces two outputs: the trip table T_{ij} as well as a vector R_i that corresponds to the rent of the TAZ. This vector corresponds to easily available real-world data. It is used as a control data set to calibrate the location choice model. Therefore, this solution does not require an estimation of a subset of T_{ij}. The advantage of this setting is that H_i, J_i and R_i can be easily updated, have a lower dimension, are more reliable and less expensive to acquire.

2. Transportation model

The task is to couple two modules that are sufficiently efficient in term of computer performance to describe large-scale systems. By using raw census data, the objective is to describe the travel demand at the highest resolution level. To be consistent, the transportation model has to meet the same requirements. This leads to network sizes that are in the range of dozens of thousands of links for metropolitan areas. Most static assignment models can handle such problems. However, as mentioned above, the temporal description is a central component of the study of travel behavior. Users have to make a trade-off between deriving utility from several activities and the travel impediments imposed by moving from one activity to the next. Taking into account this trade-off has two obvious implications. Firstly, the transportation model has to output time-dependent travel time patterns, hence the model should be dynamic. Secondly, some evaluations of activity schedules have to be integrated. Therefore the transportation model, or any co-existing module, has to model the decision of the time of usage. Note that these purposes preclude

Figure 2: Integration of location choice model and transportation model



traffic micro-simulators from the current study. Indeed, most of them are input with time-dependent trip tables (i.e. dynamic O-D matrices) and do not feature a feedback of traffic conditions on the users decision of the time of usage. Moreover, most of them cannot scale to large systems. Eventually, there is only a limited number of transportation models that can cope with these requirements: TRANSIMS, MASIM and METROPOLIS. In TRANSIMS, the beginning and the duration of activities are computed in the *Activity Generator* module. MASIM stands for a Multi-Agent SIMulation and is developed by the team of Kai Nagel at ETH Zürich. In this architecture, users are described individually through the whole simulation. The simulation of vehicle movements is performed using parallel computer resources so that it can achieve simulations of such large-scale systems faster than real-time. The scheduling of activities in MASIM is delegated to a specific module presented in this same conference (Charypar and K.Nagel, 2003). METROPOLIS (de Palma and Marchal, 2002) is used in this study. It is a dynamic traffic assignment model that embeds a departure time choice model based on the original model of Vickrey. Vickrey was the first to recognize the importance of schedule delay costs (Vickrey, 1969). Users incur schedule delays because of traffic congestion. Indeed, the theoretical finding is that schedule delay costs can be of the same order of magnitude as congestion costs. Consequently, it seems relevant to include even a simple approximation of these costs in any urban location choice model. In METROPOLIS, the travel cost $C(t)$ for a given time of usage t is the sum of the travel time cost and schedule delay cost:

$$C(t) = \alpha\tau(t) + d(t + \tau(t) - t^*) \;, \tag{2}$$

where α is the value of time, $\tau(t)$ is the travel time, $t^*$is the desired arrival time and $d(.)$ is the schedule penalty function. METROPOLIS implements a stochastic departure time choice with a continuous logit model. The time-independent consumer surplus CS_{OD} derived from traveling from O to D is given by:

$$CS_{OD} = \mu_T \ln \int_0^T \exp\left[\frac{-C_{OD}(u)}{\mu_T}\right] du \,, \tag{3}$$

where μ_T is the scale parameter of the logit model and T is the simulation period. This value is used as an input by the location choice model in order to take into account the within-day variability of travel costs. Consequently, the coupling between the location choice model and the travel model is relatively loose in the sense that the time-dependent details are hidden to the location choice model. A similar coupling is being investigated to apply METROPOLIS and URBANSIM together.

3. Location choice model

3.1 Formulation

It is assumed that the studied region is spatially discretized into n cells for which the following census data are available: H_i the number of active residents per cell and J_i the number of jobs per cell. Assume further that the total number of employees is equal to the total number of jobs N (i.e. there is market clearing for the whole city: labor supply and labor demand are at equilibrium):

$$\sum_{i=1}^{n} H_i = \sum_{j=1}^{n} J_j = N. \tag{4}$$

The objective is to derive the trip table T_{ij} from the knowledge of vectors H_i and J_j. Let us restrict ourselves to home-to-work trips. In reality, users adapt their home location to their job and conversely. At the long run equilibrium and under constant conditions, the trip distribution would verify:

$$\begin{cases} \sum_{i=1}^{n} T_{ij} = J_j \\ \sum_{j=1}^{n} T_{ij} = H_i \end{cases} . \tag{5}$$

In this restricted case, utility is only derived from two activities: working and living at a specific place. The utility for an individual working at location j and living at location i is assumed to be given by:

$$U_{ij} = CS_{ij} - r_i + w_j + \mu_R \varepsilon_{ij},$$

where CS_{ij} is the consumer surplus associated to a journey from cell i to cell j, r_i is the average rent at location i, w_j is the average wage at location j, μ_R is a scale parameter and ε_{ij} are random variables. The term r_i can also be interpreted as a (negative) residential location premium. No restrictive assumption is made on how CS_{ij} is computed, whether with a static or dynamic traffic model. The probability that an individual works at cell j and lives at cell i simultaneously is given by

$$P_{ij} = Pr\left[U_{ij} > U_{kl}, \forall k \neq i, l \neq j\right].$$

If ε_{ij} are assumed to be i.i.d. Gumbel distributions, it yields the logit formula:

$$P_{ij} = \frac{\exp\left[\frac{CS_{ij} - r_i + w_j}{\mu_R}\right]}{\sum_{k=1}^{n}\sum_{l=1}^{n} \exp\left[\frac{CS_{kl} - r_k + w_l}{\mu_R}\right]}.$$

Let us consider first the case where the job location is given. The law of conditional probability imposes that

$$P_{ij} = P\left(i \cap j\right) = P\left(i|j\right) P\left(j\right)$$

where $P\left(i|j\right) = Pr\left(h = i|w = j\right)$ and

$$P\left(j\right) = \sum_{i=1}^{n} P_{ij} = \frac{\sum_{i=1}^{n} \exp\left[\frac{CS_{ij} - r_i + w_j}{\mu_R}\right]}{\sum_{k=1}^{n}\sum_{l=1}^{n} \exp\left[\frac{CS_{kl} - r_k + w_l}{\mu_R}\right]}.$$

The conditional probability $P(i|j)$ to live in i when having a job fixed at location j is given by :

$$P(i|j) = \frac{P_{ij}}{P(j)} = \frac{\exp\left[\frac{CS_{ij}-r_i+w_j}{\mu_R}\right]}{\sum_{i=1}^{n}\exp\left[\frac{CS_{ij}-r_i+w_j}{\mu_R}\right]} = \frac{\exp\left[\frac{CS_{ij}-r_i}{\mu_R}\right]}{\sum_{k=1}^{n}\exp\left[\frac{CS_{kj}-r_k}{\mu_R}\right]}$$

Thus $P(i|j) = P(i|j)(\mathbf{r})$ and the trip table T_{ij} can be written as

$$T_{ij} = J_j P(i|j)(\mathbf{r}) \ .$$

The equilibrium condition (5) imposes that

$$\sum_{j=1}^{n} T_{ij} = \sum_{j=1}^{n} J_j P(i|j)(\mathbf{r}) = H_i \ . \tag{6}$$

Conversely, if we assume that the residential location is fixed in i, the probability to have a job in location j is given by:

$$P(j|i) = \frac{P_{ij}}{P(i)} = \frac{\exp\left[\frac{CS_{ij}+w_j}{\mu_R}\right]}{\sum_{j=1}^{n}\exp\left[\frac{CS_{ij}+w_j}{\mu_R}\right]} = P(j|i)(\mathbf{w}) \ .$$

and the trip table elements are given by $T_{ij} = H_i P(j|i)(\mathbf{w})$. The equilibrium requires that:

$$\sum_{i=1}^{n} T_{ij} = \sum_{i=1}^{n} H_i P(j|i)(\mathbf{w}) = J_j \ . \tag{7}$$

We obtain two systems of n non-linear equations (6) and (7). Obviously, rent and wages are defined up to an additive constant.

3.2 Resolution

Let $x_i = \exp\left(-\frac{r_i}{\mu_R}\right)$, $y_i = \exp\left(\frac{w_i}{\mu_R}\right)$ and $c_{ij} = \exp\left(\frac{CS_{ij}}{\mu_R}\right)$. The two systems (6) and (7) become:

$$H_i = x_i \sum_{j=1}^{n} \frac{J_j c_{ij}}{\sum_{k=1}^{n} c_{kj} x_k}$$

$$J_j = y_j \sum_{i=1}^{n} \frac{H_i c_{ij}}{\sum_{k=1}^{n} c_{ik} y_k}$$

Obviously the two systems are the same problem for the permutations $H_i \Leftrightarrow J_i$, $x_i \Leftrightarrow y_i$,$c_{ij} \Leftrightarrow c_{ji}$.

3.2.1 Special case: no transportation cost

If $c_{ij} = c_0$, then

$$H_i = x_i \frac{\sum_{j=1}^n J_j}{\sum_{k=1}^n x_k} = x_i \frac{N}{\sum_{k=1}^n x_k} .$$

The factor $K = \frac{N}{\sum_{k=1}^n x_k}$ is independent of i so that $\frac{H_i}{x_i}$ is constant. Since x_i are known up to a multiplicative factor, we can assume $K = 1$ and the solution without transportation cost is given by: $x_i^{C0} = H_i$. Similarly, we have $y_i^{C0} = J_i$.

3.2.2 Linearization

Let $f_i(\mathbf{x}) = x_i \sum_{j=1}^n \frac{J_j c_{ij}}{Z_j(\mathbf{x})} - H_i$ and $Z_j(\mathbf{x}) = \sum_{i=1}^n x_i c_{ij}$. To find the roots of $f_i(\mathbf{x})$, the Newton method is applied. Assume that $f(\mathbf{x}) \approx f\left(\mathbf{x}^{C0}\right) + D\left(\mathbf{x} - \mathbf{x}^{C0}\right)$ where D is the Jacobian matrix $D_{ij} = \frac{\partial}{\partial x_j} f_i(\mathbf{x})$. The linear system to be solved is $D\mathbf{x} = D\mathbf{x}^{C0} - f\left(\mathbf{x}^{C0}\right)$. The elements of the Jacobian matrix D are given by:

$$\begin{aligned} D_{ij} &= \frac{\partial}{\partial x_j} \left(x_i \sum_{k=1}^n \frac{J_k C_{ik}}{Z_k(\mathbf{x})} \right) \\ &= \delta_{ij} \sum_{k=1}^n \frac{J_k C_{ik}}{Z_k(\mathbf{x})} - x_i \left(\sum_{k=1}^n \frac{J_k C_{ik} C_{jk}}{Z_k^2(\mathbf{x})} \right) \end{aligned} \quad (8)$$

3.2.3 Implementation

The Newton method suffers from the disadvantage that it requires to be started sufficiently close to the solution. Preliminary experiments (not reported here) showed that starting from $\mathbf{x}^{C0}$ is not sufficient in practical cases. Therefore, iterative substitutions are performed first:

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \left(\frac{H_i}{\sum_{j=1}^n \frac{J_j c_{ij}}{Z_j(\mathbf{x}^k)}} - \mathbf{x}^k \right), \quad (9)$$

where $\mathbf{x}^k$ refers to the solution at iteration k and $\alpha \in]0, 1]$. The complete algorithm is shown below.

Empirical values are $\epsilon_1 = 10^{-2}$, $\epsilon_2 = 10^{-5}$ and $\alpha = 0.5$. Typical experiments require about a hundred iterative substitutions (loop 2-3) and less than ten iterations of the Newton method (loop 4-5-6). Step 5 is performed using the general routine DGESV from the LAPACK library (LAPACK, 2003). This method solves $A\mathbf{x} = B$ in the general case without any assumption on the properties of A. Table 1 presents the numerical performances for random data sets. All the computation times reported in this paper correspond to a PC equipped with a Pentium 4 CPU clocked a 2.5Ghz. The computation times are roughly proportional to n^3. This is consistent with the fact that the computation of the Jacobian is of order n^3, as well as the LU factorization performed by DGESV.

Algorithm 1 Solving the location choice model

1. Set $k = 0$. Initialize the rent vector $x_i^0 = x_i^{C0} = H_i$.
2. Compute $\mathbf{x}^{k+1}$ using (9).
3. If $\frac{\|\mathbf{x}^{k+1}-\mathbf{x}^k\|}{\|\mathbf{x}^k\|} > \epsilon_1$, go to step 2.
4. Compute D_{ij}^k using (8).
5. Solve $D^k\mathbf{x}^{k+1} = D^k\mathbf{x}^k - f\left(\mathbf{x}^k\right)$.
6. If $\frac{\|\mathbf{x}^{k+1}-\mathbf{x}^k\|}{\|\mathbf{x}^k\|} > \epsilon_2$, go to step 4.

Table 1: Numerical performance of the location choice model.

n	CPU [s]	RAM [Mb]
500	20	<5
1,000	175	23
2,000	1850	92

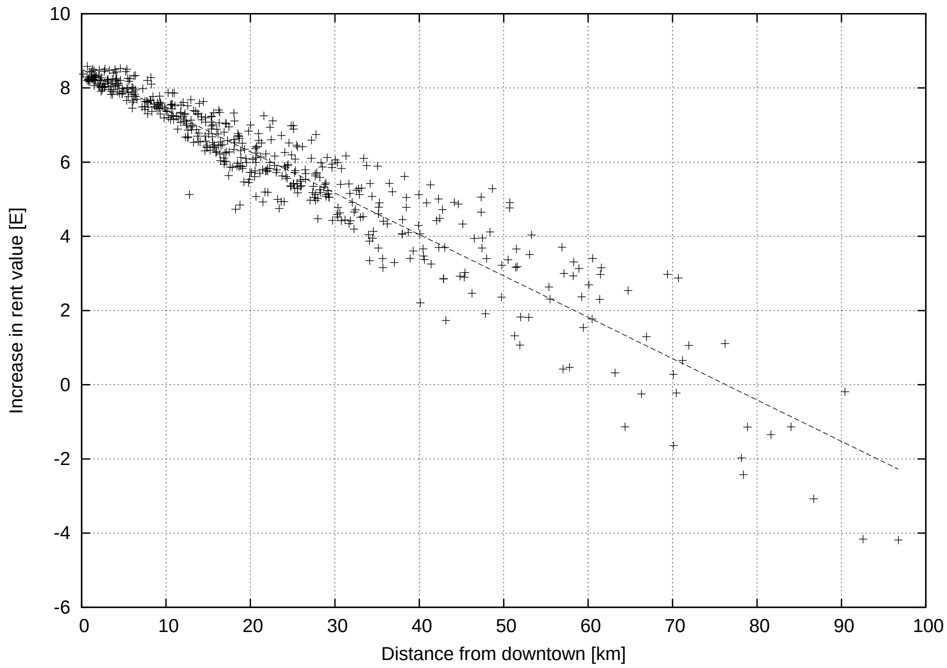
4. Applications to medium-scale systems

A detailed application of the above model was performed on the Paris area in a recent paper (de Palma and Marchal, 2003). The METROPOLIS transportation model was calibrated using external sources such as surveys and traffic counts. The Parisian regional database covers about 20,000 square kilometers and was provided by a local agency, IAURIF. The coded network is made of 18,000 links and is not fully disaggregated (only the main roads are coded). The origin and destination locations are grouped into 514 TAZ. The calibration of the location choice model consists only in the adjustment of the parameter μ_R to fit rent patterns r_i with actual rent values. In this study the calibration was carried out in reverse order. An external trip table T_{ij} calibrated with traffic counts was used to adjust μ_R. Then, the rent patterns produced by the model were compared to actual values collected for the Paris area. The calibration process yielded μ_R=1.75$ while the average car travel cost for a commuting journey in the morning was 4.91$. The Paris area is strongly CBD-oriented. Therefore, it is interesting to plot the rent gradient as a function of the distance from downtown Paris. Figure 3 shows that the gradient is about 0.11$ per km. This value can be interpreted as the increment in rent value due to the presence of travel costs. At first sight, this is a rather small figure, but it corresponds to the rent gradient due to a single trip. If we assume 50 trips in a month on average, this yields a rent gradient of about 5.5$ per month per kilometer. Comparison with real-world values (not presented here) revealed that the location choice model captured about half of the rent discrepancies between downtown and the suburbs. However, the rent in the CBD was underestimated.

The resolution of the location choice model is rather fast for a problem of this size (see Table 1) compared to the computation of the transportation model. METROPOLIS requires typically two hours of CPU time to compute the traffic equilibrium for this problem. It corresponds to about 50 iterations (e.g. morning peaks). An important issue is the convergence of the coupling of the models (see Figure 2). The following iterative procedure was applied:

1. Year 0. $c_{ij} = c_0$. Compute T_{ij}^0 and r_i^0 using the location choice model.

Figure 3: Rent increment due to single trip cost and linear regression.



2. Year 0. Apply METROPOLIS, compute c^1_{ij}.
3. Year 1. Compute T^1_{ij} and r^1_i using the location choice model and c^1_{ij}.
4. Year 1. Apply METROPOLIS, compute c^2_{ij}.

The relative standard deviation between years 0 and 1 for the rent patterns is 106%, then it drops to 5% between years 1 and 2, then 4% between years 2 and 3. Two steps are sufficient in practice to obtain a stable trip distribution. The traffic equilibrium can exhibit some differences between successive steps (not presented here), such as peak congestion being trade-off for schedule delay. However, the computation of the consumer surplus using (3) with $\mu_T = 1\$$ tends to smooth these variations.

5. Applications to large-scale systems

In the previous section, it has been shown that the location choice model can be used to produce realistic trip tables for the primary activity but that the algorithm is poorly scalable. If the density of active residents and jobs is provided using census raster data instead of zones, the size of the system will put the problem out of reach of the previous algorithm (e.g. a 10km by 10km urban area corresponds to 10,000 cells). The ultimate goal is to develop a methodology that will scale beyond the limit of a few thousands cells and that avoid, if possible, a complete enumeration of all the possible alternatives. Several potential tracks can be examined to solve this limitations.

1. The Newton method is not the most efficient way to solve the non-linear system. In particular, it does not exploit the property that it is a fixed point problem. Generalization of secant methods have been proposed recently (Crittin and Bierlaire, 2003) that can handle problems of size 100k. Experiments have not yet been conducted to compare computing times for this specific problem.

2. An agent-based optimization simulation is under development where agents have a limited choice set (several dozens of cells at most). From this set they select their primary and secondary activities. An update process circulates cell information via social connections between agents. Preliminary results (not reported here) indicates that this process might be slower than a full enumeration but that it has the advantage to scale independently of the size of the grid. Moreover, it can potentially capture the dynamics of the modification of behaviors due to external changes.

3. Cells can be automatically clustered into TAZ. Obviously this is a second-best solution since the goal is to get rid of the ambiguous TAZ description. Since different grouping criteria produce different TAZ patterns, this solution is not completely suitable. Nevertheless, it can provide, in a first approximation, a set of reasonable initial conditions for the multi-agent simulations. This can speed up the convergence by starting the simulation with a plausible choice set for each agent.

5.1 Cell clustering

Ideally, the automatic clustering of cells into TAZ should be adaptative and should only depend on the data and on the structure of the problem. It has been noted before than the rent values computed by the location choice model are defined up to an additive constant. These results only make sense when compared to the solution without travel costs x^{C0}. Recall that $r_i^{C0} = -\mu \ln x_i^{CO} = -\mu \ln H_i$. This implies that if a TAZ is divided into two children TAZ with the same amount of residents $\frac{H_i}{2}$, the rent without travel costs is increased by a factor $\mu \ln 2$. (This can be seen as an instance of the IIA problem of the logit model.) One one hand, it is suitable to choose a clustering rule such that the location choice model produces the widest spectrum of rent values. On the other hand, it is not suitable to introduce discrepancies that are artifacts in the rent patterns. Therefore, the clustering rules are selected as follows:

- group only neighboring cells,
- group cells so that TAZ have the same amount of residents H_i.

If all TAZ have the same amount of residents, they should have all the same rent value in the case without travel costs and we can assume $r^{C0} = 0$ for all the TAZ. Note that this method can group cells with quite different properties in term of residents, jobs and travel costs. However this rule spares us the definition of a similarity function for the cells. The first rule prevents TAZ to be grouped into non-connected sets and allows to use graph partitioning techniques to implement the clustering.

The implementation of the clustering is performed using a graph partitioning package (METIS, 2003). Cells are assumed to be vertices of a lattice where each node is connected to the north, east, south and west to other cells. Edges have no weight and vertices have weights equal to H_i. METIS can compute a graph partition that ensures an equal distribution of the weights in each cluster. Figure 4 presents this encoding. Empty cells can be connected at no cost, to allow isolated non-empty cells to be merged with neighboring TAZ. Eventually, the center of gravity P_g of the TAZ is computed to determine its location: $P_g = \frac{\sum_{cells} p_g H_i}{\sum_{cells} H_i}$.

Figure 4: Graph partitioning of the cells

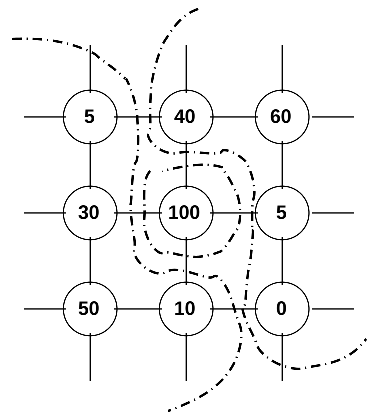


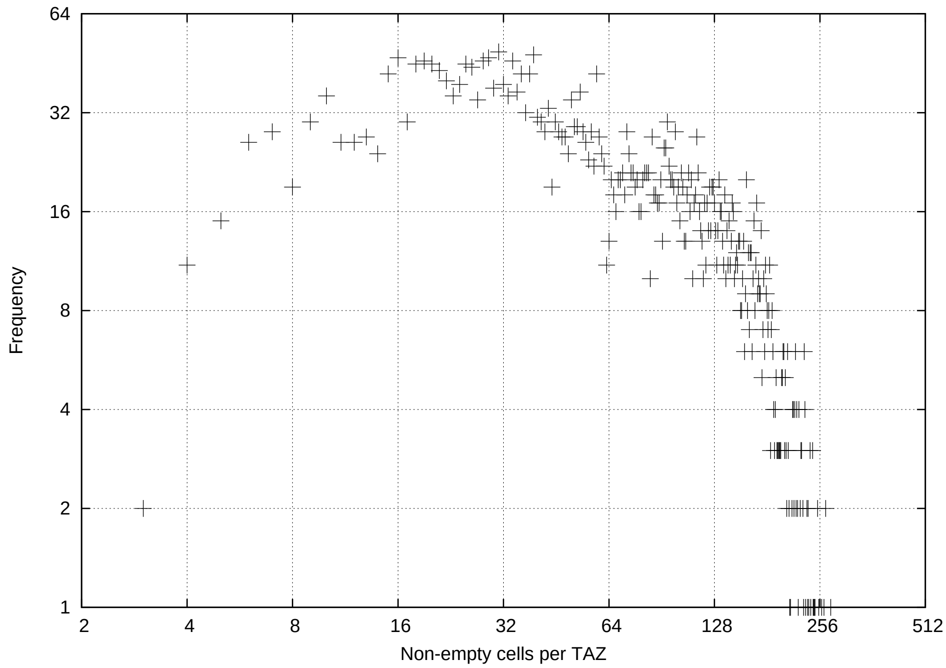
Table 2: Efficiency of the clustering method

	Zurich	Switzerland
Size [kmxkm]	80.3x70.1	345.8x219.4
#cells [k]	563	7586
#non-empty cells [k]	81	316
#TAZ	1,000	4,000
$< H_i >$	1195.2	1025.78
σ_{H_i}	23.65	22.12
CPU [s]	19	278
RAM [Mb]	57	961

5.2 Results

The data set for the region of Zurich consists of a subset of the Swiss census and covers about 563,000 cells (only 81,000 have data). The corresponding coded network comes from the NavTech company and contains more than 76,000 links. Figure 6 displays the clustering into 1,000 TAZ of the Zurich area. It can be seen that more TAZ are produced where the density of residents is higher, close to Zurich downtown.The efficiency of the graph partitioning is presented in Table 2. Note that the efficiency of the METIS routine could be considerably improved by building a graph that handles empty cells in a more sophisticated manner since, in this application, most of the vertices have null weights. Nevertheless, the computing times are acceptable even for large-scale systems. Figure 5 presents the distribution of the number of non-empty census cells per TAZ for the whole Switzerland data set clustered with the same technique. Figure 6 presents the monthly rent pattern for the Zurich area. Rents are given in CHF and are computed assuming 50 trips a month. Rents have been adjusted to a minimum of 1,000 CHF to be comparable with real values.

Figure 5: Clustering of the whole Switzerland census data



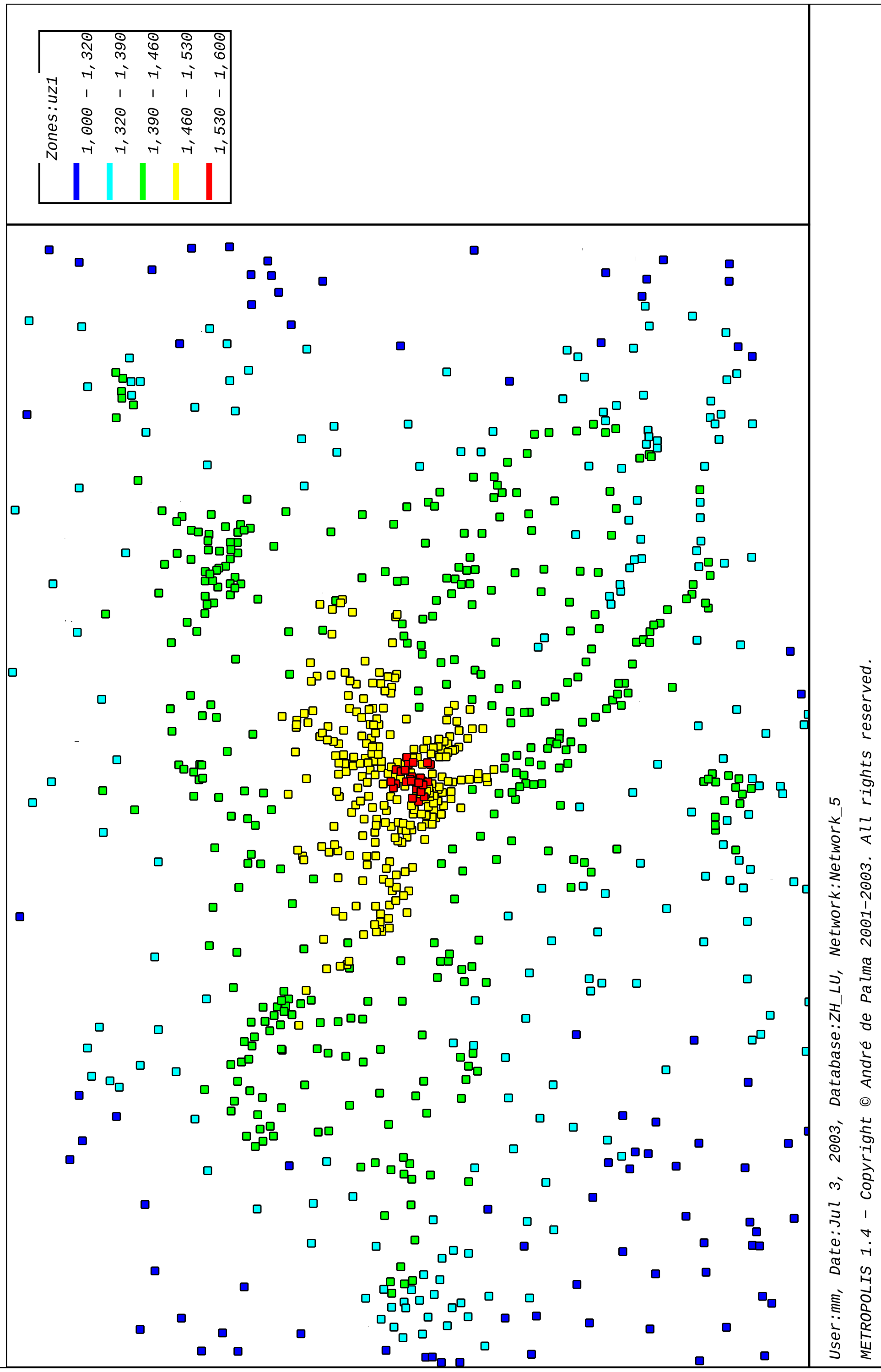
6. Conclusions

This paper proposed an efficient solution to provide large-scale simulations of urban areas with trip distribution in a more consistent manner than the four-step planning scheme. In particular, the proposed model computes in a few steps a trip distribution that depends indirectly on within-day travel time variability as well as on activity schedule constraints. The deliberate choice to keep a simple description (home-to-work trips, car traffic only, linear schedule delay) allows to design a fast procedure that uses raw census data in an straightforward way. Nevertheless, some scalability problems remain. A workaround has been elaborated using cell clustering. For large-scale systems with hundred of thousands of cells, further research directions have been suggested, such as multi-agent simulations and efficient solvers for large-scale fixed point problems.

Acknowledgments

The author expresses his thanks to Prof. K. Axhausen for providing the data for the Zurich area and to Prof. A. de Palma and K. Nagel for their helpful discussions. Computer resources for the simulations were provided by the Computational Laboratory (CoLab) at ETHZ.

Figure 6: Monthly rent patterns (in CHF) and cell clustering for the Zurich area.



References

Ben-Akiva, M. and J. Bowman (1998) Activity based travel demand model systems, *Equilibrium and Advanced Transportation Modelling, P. Marcotte and S. Nguyen eds., Kluwer Academics*, 27–46.

Charypar, D. and K.Nagel (2003) Generating complete daily activity plans with genetic algorithms, *Paper to be presented at the 10th International Conference on Travel Behaviour Research*.

Crittin, F. and M. Bierlaire (2003) A generalization of secant methods for solving nonlinear systems of equations, *Paper presented at the third Swiss Transportation Research Conference, Ascona*.

de Palma, A. and F. Marchal (2002) Real cases applications of the fully dynamic METROPOLIS toolbox: an advocacy for global large-scale mesoscopic transportation systems, *Networks and Spatial Economics 2(4)*, 347–369.

de Palma, A. and F. Marchal (2003) Integration of dynamic traffic models and land-use forecasting with an application to the Paris area, *Paper submitted to Regional Science and Urban Economics*.

LAPACK (2003) Linear Algebra PACKage, URL `www.netlib.org/lapack`.

METIS (2003) Family of multilevel partitioning algorithms, URL `www-users.cs.umn.edu/ karypis/metis`.

Ortuzar, J. D. and L. G. Willumsen (1994) *Modelling Transport*, John Wiley and Sons, New York, second edn.

Raney, B., N. Cetin, A. Völlmy, M. Vrtic, K. Axhausen and K. Nagel (2003) An agent-based microsimulation model of Swiss travel: First results, *Networks and Spatial Economics 3(1)*, 23–41.

TRANSIMS (2003) TRansportation ANalysis and SIMulation Systems, URL `transims.tsasa.lanl.gov`.

Vickrey, W. (1969) Congestion theory and transport investment, *American Economic Review (Papers and Proceedings) 59*, 251–261.

Waddell, P., A. Borning, M. Noth, N. Freier, M. Becke and G. Ulfarsson (2003) Microsimulation of urban development and location choices: Design and implementation of URBANSIM, *Networks and Spatial Economics 3(1)*, 43–67.