Calibration of Choice Model Parameters in a Transport Scenario with Heterogeneous Traffic Conditions and Income Dependency

A. Agarwal and D. Ziemke and K. Nagel

Transport Systems Planning and Transport Telematics
Technische Universität Berlin, Berlin, 10587, Germany
Tel. +49 (0) 30-314-23308, Tel. +49 (0) 30-314-23308

Word count: ∼ 6600

ABSTRACT
By raising the issue of data requirements for the purpose of modal development, validation and application, this study proposes an approach to calibrate choice model parameters in heterogeneous traffic condition using minimal empirical data. For this, a real-world scenario of Patna, India is chosen. For the calibration, a Bayesian framework based calibration technique (CaDyTS: Calibration of Dynamic Traffic Simulations) is used. Commonly available, mode-specific, hourly-classified traffic counts are used to generate full day plans of agents and their initially unknown activity locations. While the proposed approach implements location choice implicitly, the approach can be applied to a variety of other problems. Further, the effect of household income is included in the utility function to incorporate the effect of income in the decision making process of individual travelers and to filter out inconsistencies in the daily plans, which originate from the survey data.

1. Introduction

In a transportation system, a wide variety of data (e.g. network data, socio-economic data) is required for the purpose of model development, validation and application. The aim of such models is to simulate and analyze travel demand, and test the policies, which can help transport planners to understand the decision making process of individual travelers. A model should be causal, flexible, transferable, efficient, and sensitive to policy objectives (Domencich and McFadden 1996). Most travel demand models minimally require information about the trip origin, trip destination, and trip mode. The information about origin and destination (OD) can come in different forms and at different level of aggregation, e.g. as an OD matrix, as daily plans, etc. The traditional way to estimate the OD matrix relies on roadside or household surveys, which are, however, error-prone and likely to be biased (Kuwahara and Sullivan 1987; Groves 2006). As an alternative, there are several approaches to estimate the OD ma-

---

CONTACT A. Agarwal. Email: agarwal@vsp.tu-berlin.de

trix using traffic counts (e.g. see Bell (1983); Cascetta, Inaudi, and Marquis (1993); van Zuylen and Willumsen (1980)).

Given the origin-destination information of an area, static traffic assignment (STA) provides the traffic flow on each highway for every time bin. Dynamic traffic assignment (DTA) is a generalization of STA, which provides time-dependent traffic flow on each highway segment (Szeto and Wong 2012). From the development perspective, DTA models can be classified in two categories, analytical (Chen and Hsueh 1998) and simulation-based models. The former are often preferred for small networks whereas the latter are preferred for realistic networks of large urban agglomeration and for microscopic traffic flow characteristics (Szeto and Wong 2012; Bliemer 2007). In the context of the application of such models to large urban transportation networks, at least two problems become apparent: a) microscopic modeling is computationally expensive and b) data requirements are high. Mainly based on the underlying traffic flow model, DTA models can be classified as physical-queue models (Szeto 2008; Szeto, Jiang, and Sumalee 2011) and non-physical-queue models (Lam and Huang 1995; Szeto and Wong 2012). One such physical-queue model (Gawron 1998; Cetin, Burri, and Nagel 2003) is embedded in the activity-based, multi-agent transport simulation framework MATSim (Horni, Nagel, and Axhausen 2016). Due to its simplicity, it is able to handle large urban transportation networks (Balmer et al. 2008) and still resembles to Newell's simplified kinematic wave model (Agarwal, Lämmel, and Nagel 2016; Agarwal, Lämmel, and Nagel 2017). The aforementioned problem, regarding the resource-intensive models can be managed by such fast traffic flow models.

Traditionally, in order to gather the required data, different types of data collection techniques are used, which are either manual or automatic. Such approaches include mid-block traffic count surveys, spot-speed surveys, origin-destination surveys, household surveys etc. (Currin 2012). The use of mid-block traffic counts survey is popular in India for various purposes. However, this information is not sufficient to simulate the travel demand for an urban scenario in order to understand the behavior of individual travelers. The complexity rises if traffic streams are populated with different vehicle types, which is very common in most developing economies. In this direction, this study proposes an approach to calibrate travel demand in heterogeneous traffic conditions using minimal empirical data.

In contrast to traditional data collection techniques, several studies apply alternative approaches to derive and validate travel demand. Detailed surveys to collect the data (e.g. household surveys), which require origin and destination information, trip modes, trip purposes, start times, end times etc. are often associated with high non-responses and misreporting rates (Zimowski et al. 1997; Wolf 2000). Traffic data collection based on manual or automated traffic counts is usually easier to manage. With the recent technological advances, new approaches are presented, which make use of GPS (Geographical Positioning System) technology in traditional travel surveys, which is likely to improve the quality and robustness of the data (Wolf 2000; Chung and Shalaby 2005; Shen and Stopher 2014). With a web-based survey, it is shown that innovative sources to collect travel data is gaining popularity as well as acceptance (Lee, Sener, and Mullins III 2016). GPS data is also used to study the decision making process of cyclists (Hood, Sall, and Charlton 2011), to measure and visualize space-time congestion patterns (Stipancic et al. 2017). Similarly, in the last couple of years, several other studies proposed different approaches to collect data using CDR (call detail records) from smart-phones (Iqbal et al. 2014; Chen and Bierlaire 2014). A simulation-based approach to construct all-day trip chains using mobile phone data

is proposed by Zilske and Nagel (2015), which reduces spatio-temporal uncertainties.[1]

From the above background, main focus of the study is to explore an approach to use traditional or modern data effectively for the purpose of constructing synthetic activity-trip chains of individual travelers which is essential in activity-based simulations. This study proposes an approach to construct trip diaries in heterogeneous traffic conditions using hourly classified mid-block traffic counts. That is, In addition to this, household income levels are incorporated in the utility function to understand the choices of travelers. For this, a real-world scenario of Patna, India, is considered. The data for the scenario is taken from the Comprehensive Mobility Plan (CMP) for Patna (TRIPP, iTrans, and VKS 2009). A few inconsistencies in the survey data are observed, which are likely to occur in other scenarios as well. Some of these inconsistencies are repaired in the scenario. The remainder of the paper is structured as follows. Section 2 illustrates the calibration process, Section 3 exhibits the travel demand for the scenario and construction of an income-dependent utility function. Calibration results are presented and discussed in Section 4. The study is concluded in the Section 5.

2. Calibration procedure

In this study, the multi-agent based transport simulation framework MATSim is used (see Section 2.1), which is able to handle large-scale scenarios because of its fast network loading algorithm and ability to handle mixed traffic conditions (Agarwal et al. 2015; Agarwal and Lämmel 2016). Together with this, the calibrator CaDyTS ('**Ca**libration of **Dy**namic **T**raffic **S**imulations'; see Section 2.2) is used. It has been used previously to adjust traffic demand of car traffic (Flötteröd, Chen, and Nagel 2011; Ziemke, Nagel, and Bhat 2015) and to calibrate the travel demand for public transit (Moyo Oliveros and Nagel 2012). It has also been applied to solve the problem of location choice (Ziemke, Nagel, and Bhat 2015), which was applied in the creation of an open scenario for Berlin (Ziemke and Nagel 2017). In these approaches, however, CaDyTS was used for homogeneous traffic conditions, while the present study extends the approach for heterogeneous traffic conditions.

2.1. Travel Simulator: MATSim

In this study, the MATSim transport simulation framework (Horni, Nagel, and Axhausen 2016) is used for all simulation experiments. The minimal inputs for a simulation run are the physical boundary conditions (i.e. the road network), daily plans of individual travelers and scenario-specific parameters. The network loading algorithm of MATSim is embedded to an iterative cycle in which every individual traveler is considered as an agent. The cycle consists of following three parts:

(1) Plans execution: In this step, the plans of all individual travelers are executed simultaneously on the network using a mobility simulation. In this study, a time-step-based queue simulation approach (Gawron 1998; Simon, Esser, and Nagel 1999) is used. This can also simulate heterogeneous traffic conditions realistically (Agarwal et al. 2015; Agarwal and Lämmel 2016; Agarwal 2017; Agarwal, Lämmel, and Nagel 2017).

---

[1] Refer to Rieser-Schüssler (2012); Lee, Sener, and Mullins III (2016); Barmpounakis et al. (2017) for more details about the modern data collections approaches, data sources and examples.

(2) Plans evaluation: The executed plans are evaluated using a utility (scoring) function. In this study, the default 'Charypar-Nagel' scoring function (Charypar and Nagel 2005) is used and further modified to include the effect of household income (see Section 3.2.2).

(3) Re-planning: This step is composed of two parts i.e. innovation and plan selection. A new plan is generated for some agents by modifying an existing plan's attribute (departure time, route, mode etc.) using so-called innovative strategies. The new plan is executed in the next iteration. Innovation is used until fixed number of iterations. The old plans are kept in the agents' memories; the worst plan is removed from choice set if maximum number of plans in the choice set of a person is reached. Agents which do not undergo innovation, select a plan from their choice set using so-called non-innovative strategies (i.e., plan selection).

The above steps are repeated in an iterative process. Finally, a number of additional iterations are run only with non-innovative strategies which finally results in stabilized simulation outputs.

2.2. *Calibrator: CaDyTS*

In an activity-based simulation framework, traffic counts are insufficient to generate whole day plans of individual travelers. To address this issue, a calibrator called 'CaDyTS' is used ('**Ca**libration of **Dy**namic **T**raffic **S**imulations'; Flötteröd, Bierlaire, and Nagel 2011; Flötteröd 2010), which is based within a Bayesian framework. Together with simulation framework, this is integrated to the utility function such that probability of selecting a plan i from the j plans is given by Equation (1). In this, y_{lt} and q_{lt} are the measurement and simulation values for spatial location l and time bin t. σ_{lt}^2 is variance of measurement. V_i is the utility of the plan and ω is weight parameter for correction ΔV_{lt} (Equation (2)).

$$P(i|y) = \frac{exp(V_i + \omega \cdot \sum_{lt} \Delta V_{lt})}{\sum_j exp(V_i + \omega \cdot \sum_{lt} \Delta V_{lt})} \tag{1}$$

$$\Delta V_{lt} = \frac{y_{lt} - q_{lt}}{\sigma_{lt}^2} \tag{2}$$

In this study, hourly classified traffic counts are available, which are used to generate whole day plan for the travelers. From the Equations (1) and (2), one can observe that a plan, in which, an agent traverses a link whose simulated counts are underestimated, is more likely to be chosen. For heterogeneous traffic conditions, Equation (2) is modified as shown in Equation (3); where m is the mode for which measured traffic counts at link l, time bin t are available:

$$\Delta V_{ltm} = \frac{y_{ltm} - q_{ltm}}{\sigma_{ltm}^2} \tag{3}$$

Revisiting Equations (1) and (3), it can be observed that, if the choice set of an agent contains plans with different modes, the correction is likely to fix the modal share as well. In this study, CaDyTS is used to generate full day plans of agents and its

initially unknown activity locations. The choices for the different activity locations are provided by creating multiple plans corresponding to each plausible activity location (see Figure 1). The calibration approach can be applied to a variety of problems.

3. Real-world case study: Patna, India

This section exhibits the set-up for a real-world scenario of Patna, India. The road network, survey locations, and the land-use patterns of Patna are shown in Figure 1 (Agarwal 2017).

3.1. *Travel Demand*

The travel demand of the region is categorized in two groups, urban and external travel demand.

3.1.1. *Urban travel demand*

Urban travel demand is generated directly from a trip diary survey (TRIPP, iTrans, and VKS 2009). Table 1 shows the modal income statistics for households of Patna city. This data is evaluated from individual monthly income form trip diaries.[2] Car is predominantly used by high income persons whereas motorbike is used by mid to high income persons. Bicycle and walk trips are limited to low income households. Trip diaries result in 13,278 records which represent approximately 1% sample of all trips. Every such record is translated into one agent with one plan. In absence of other data, for each plan two trips are generated, one 'to work/education/social/other' and one 'back home'. This is somewhat similar to generating an AM peak and a PM peak origin-destination-matrix. In order to get significant number of plans for commuters and through traffic in various categories (see Section 3.1.2 and Appendix A), the data is expanded to a 10% sample. Therefore, urban plans are cloned as follows:

(a) The origin and destination zones of each trip are known from household survey data. For every person, a random point is taken from the origin and destination zones i.e. all cloned persons are likely to originate and terminate on different links.

(b) Same travel mode is assumed for all cloned persons to maintain the modal share distribution from survey data.

(c) A trip starts immediately after ending an activity (e.g. home, work, education etc.). The activity end time for each activity is randomized within a plausible range depending on the trip purpose. For instance, a person departs between 08:00 to 09:30 for work, between 6:30 to 08:30 for education etc. Typical durations for home, work, education, social and other activities are assumed as 12, 8, 7, 5, 5 h respectively.

---

[2] Parts of the data in the household survey were unavailable (e.g. missing trips for few zones, missing households income for few persons etc.); for such cases the required data were imputed randomly based on other available data (e.g. trip distribution, income distribution etc.) in the Patna CMP (see Ch. 5 in Agarwal 2012, for further details about the imputation of missing trips).

Table 1. Average income (₹/month) statistics for Patna city; data is generated from trip diaries (TRIPP, iTrans, and VKS 2009).

travel mode	number of persons	mean income	median income
bicycle	3878	5903.24	4000.0
car	526	13482.41	20000.0
motorbike	2668	10341.26	6250.0
PT	3527	8343.99	4000.0
walk	2679	6383.35	4000.0
all modes	13278	7840.43	4000.0

(d) Every person has unique identifier and all cloned persons have different plan attributes (e.g. location of trip origin/destination, trip start time etc.). Thus, later in the simulation, every person is considered individually.

3.1.2. External travel demand

The *external* travel demand is further classified into through traffic and commuters. The former is the traffic which passes through Patna and consists of at most one trip per day, whereas the latter consists of agents who commute between Patna and nearby areas, and have 2 trips in their plans. To include the congestion effect of external traffic in the activity-based transport simulation framework, the whole day plans of the external traffic are required. These are generated as follows.

(1) The Patna CMP provides hourly classified counts for 7 outer cordon stations (see Figure 1) in both directions and directional split factors (see Appendix A). The directional split provides the share of commuters and through traffic from each counting station.

(2) For through traffic, an OD matrix is given, which provides the origins and destinations (see Table A3). In absence of additional information, the OD weights from the matrix are used for all modes (bicycle, car, motorbike and truck) and in all time bins; this provides the mode and departure times for the trips.[3] Consequently, a 10% sample is created from the counts such that each through traffic plan has one trip only.

(3) For commuters, exact locations of the trip destinations are initially unknown. They are calibrated in this study based on the given traffic counts in a similar way as done by Ziemke, Nagel, and Bhat (2015) for car traffic. A few potential activity locations are identified based on the land-use pattern (see Figure 1). A random point inside any of these probable activity location areas is taken as the trip destination. Thus, for every agent, 5 plans are generated corresponding to each plausible destination and added to the choice set of the agent. From Equation (1) recall that a plan is favored if the agent travels via one of the counting stations that is underestimated in the simulation. In other words, within the simulation framework, location choice is available to the agents, similar to OD matrix estimation in trip based models (e.g. Bell 1983), but estimating the location for the outgoing and the returning trip together.

---

[3] Refer to Appendix A for more details about the input data for external travel demand, steps to estimate the external trip counts, directional split and OD matrix for through traffic. This data is taken from Patna CMP (TRIPP, iTrans, and VKS 2009).

Table 2. Modal attributes for Patna scenario.

	bicycle	car	motorbike	truck	PT	walk
Speed (km/h)	15	60	60	30	20	5
PCU	0.15	1	0.15	3	–	–

Table 3. Values of time and vehicle operating costs (IRC:SP:30 2009).

travel mode	vehicle operating costs (USDct/km)	value of time (USDct/h)
car	3.75	93.84
motorbike	1.55	48.05
PT	–	59.31

3.2. Scenario preparation

The calibration of the scenario is performed for the following reasons.

(a) Trip destinations (activity locations) of the commuters are unknown.

(b) A few trip diaries do not have mode and income information which is randomly assigned based on the income-dependent modal distribution from Patna CMP (see Section 3.1.1).

(c) A few trip diaries are inconsistent (see Figure 3(a)). For instance, i) persons from very low income group (8-11 USD/month) make trips by car, ii) persons from high income group make 10 km long trips using bicycle or walk modes. Such situations are very unlikely and assumed as reporting errors.

(d) The Patna CMP does not provide any utility parameters. As a starting point utility parameters are taken from IRC:SP:30 (2009) as shown in Table 3; these parameters are not related to the CMP survey. Other elements of a mode-choice utility function, such as alternative (or mode) specific constants (ASCs) for all modes or marginal utility of distances, are unknown and need to be found from calibration.

3.2.1. Travel modes

In this study, car, motorbike, bicycle, and truck modes are physically simulated on network (so called main modes or congested modes), whereas walk and public transit (PT) are teleported between origin and destination (so-called uncongested or teleported modes). The main difference between the two is that main modes consume flow and storage capacities on the link and thus affect the route choice decision making process of the individual travelers. Table 2 provides the maximum speeds for all modes and PCU (passenger car unit) for congested modes. In the traffic mix, shares of bicycle and motorbike modes are high, therefore, the PCU of bicycle and motorbike is assumed as 0.15 (Chandra and Sikdar 2000).

3.2.2. Utility function

3.2.2.1. Utility parameters. To evaluate a plan, a scoring function is used which requires explicit values for utility parameters. In order to determine the utility parameters, the value of time and vehicle operating costs is taken from IRC:SP:30 (2009) and converted to USD[4] for a common interpretation (see Table 3). The average trip cost per km for PT is taken from Kumar, Baus, and Maitra (2004) and shown in Equa-

---

[4]1 USD ≈ 66.6 ₹. Exchange rate on 8 June 2016.

tion (4). The value are on the lower side, however, seems appropriate due to significant share of low cost 'tuk-tuks' in Patna.

$$\text{PT trip costs[USD]} = \begin{cases} 0.045, & \text{if } d \leq 4 \text{ km} \\ 0.045 + (d-4) \cdot 0.0047, & \text{if } d > 4 \text{ km} \end{cases} \tag{4}$$

3.2.2.2. Dependency on household income. In general, the value of time is the opportunity cost of time an individual traveler spends on the trip; this is highly dependent on the income level of individual. In order to incorporate the high income differentiation across different modes, the perception of income is added to behavioral decision making process of individual as follows:

(1) **Utility of traveling:** The utility of traveling is given by:

$$S_{trav,mode} = C_{mode} + \tilde{\beta}_{trav,mode} \cdot t_{trav} + (\beta_{d,mode} + \beta_m \cdot \gamma_{d,mode}) \cdot d_{trav} \tag{5}$$

where C_{mode} is ASC for mode $mode$, $\tilde{\beta}_{trav,mode}$ is the effective (see below) marginal utility of time spent traveling (normally negative or zero), $\beta_{d,mode}$ is marginal utility of distance (normally negative or zero), β_m is marginal utility of money (normally positive) and $\gamma_{d,mode}$ is mode-specific monetary distance rate (normally negative or zero). t_{trav} and d_{trav} is travel time and travel distance between two activity locations.

(2) **Marginal utility of traveling:**
  a) As is common (e.g. Franklin 2006), it is assumed that the income-dependent marginal utility of money $(\beta_{m,j})$ of person j is indirectly proportional to this person's income y_j:

$$\beta_{m,j} = \frac{\bar{y}}{y_j} \frac{\text{util}}{\text{USD}}$$

  where $\bar{y}$ is the median income for all individuals.
  b) The value of travel time savings (VTTS) is related to Equation (5) in the usual way as $-\beta_m/\tilde{\beta}_{trav}$, e.g. for car as

$$\text{VTTS}_{car} \overset{!}{=} \frac{-\tilde{\beta}_{trav,car}}{\beta_m} \ . \tag{6}$$

  It is now plausible to assume that the car VTTS values from Table 3 were obtained from people who actually used car, i.e. those with higher income. Equation (6) thus becomes

$$\text{VTTS}_{car} \overset{!}{=} \frac{-\tilde{\beta}_{trav,car}}{\beta_{m,\text{highIncome}}} \ . \tag{7}$$

  Together with

$$\beta_{m,\text{highIncome}} = \frac{\bar{y}}{y_{\text{highIncome}}} \frac{\text{util}}{\text{USD}}$$

where it is thus assumed that the car users have a 'typical' income of $y_{\text{highIncome}}$, and after rearrangement, Equation (7) becomes

$$\widetilde{\beta}_{trav,car} = -\text{VTTS}_{car} \cdot \frac{\bar{y}}{y_{\text{highIncome}}} \frac{\text{util}}{\text{USD}} = -0.9384 \frac{\text{USD}}{\text{h}} \cdot \frac{4000}{20000} \frac{\text{util}}{\text{USD}} = -0.19 \frac{\text{util}}{\text{h}},$$

where the number values are now taken from Table 3, and the income values from Table 1; note that a conversion of the Rupee values into USD is not necessary because of the division.

c) Similarly, for motorbike and PT, the marginal utility of traveling will be:

$$\widetilde{\beta}_{trav,mb} = -0.4805 \frac{\text{USD}}{\text{h}} \cdot \frac{4000}{6250} \frac{\text{util}}{\text{USD}} = -0.31 \frac{\text{util}}{\text{h}}$$

$$\widetilde{\beta}_{trav,PT} = -0.5931 \frac{\text{USD}}{\text{h}} \cdot \frac{4000}{4000} \frac{\text{util}}{\text{USD}} = -0.59 \frac{\text{util}}{\text{h}}$$

d) In absence of the values of time for bicycle and walk modes, (dis)utility (or disagreeability) of being (stuck) in traffic for bicycle and walk mode is assumed same as motorbike; i.e.

$$\widetilde{\beta}_{trav,bicycle} = \widetilde{\beta}_{trav,walk} = \widetilde{\beta}_{trav,mb} = -0.31 \text{ util/h}$$

These values now plausibly express that in terms of marginal utility of time spent traveling, car is the most favorable of all available modes, and PT the least favorable. The fact that the VTTS of car in Table 3 comes out as the one with the *highest* willingness-to-pay to shorten its duration is explained by the higher income of car users, and not as a general inconvenience of car.

(3) **Utility of performing an activity:** Considering the marginal utility of time as a resource, a unit reduction in travel time (Δt) would not only save the direct (dis)utility of travel $\beta_{trav} \cdot \Delta t$ but also increase the score by the utility of time as a resource, which approximately is $\beta_{dur} \cdot \Delta t$ (Kickhöfer and Nagel 2016). The latter is the opportunity cost of time gained by performing the activities for the saved time (Δt). This results in

$$\widetilde{\beta}_{trav,mode} = \beta_{trav,mode} - \beta_{dur}$$

where the sign convention is such that the parameter β_{dur} is typically positive, $-\beta_{dur}$ in consequence negative, and $\beta_{trav,mode}$ denotes *additional* inconvenience of the mode over 'doing nothing'. Following Kickhöfer and Nagel (2016), the value of the marginal utility of performing an activity (β_{dur}) is taken as the negative of that marginal utility of traveling across all the considered modes that is closest to zero (here thus $\beta_{dur} = -\widetilde{\beta}_{trav,car} = 0.19$ util/h), and the corresponding direct marginal utility, $\beta_{trav,car}$, is set to zero. All other direct marginal utilities of traveling are set relative to this value, i.e.

$$\beta_{trav,mode} = 0.19 \text{ util/h} - \widetilde{\beta}_{trav,mode}$$

In words: The marginal disutility of each mode is decomposed into a 'base'

Table 4. Utility parameters converted to MATSim format.

travel mode	bicycle	car	motorbike	PT	walk
monetary distance rate (γ_d) [USD/m]	–	$-3.7 \cdot 10^{-5}$	$-1.6 \cdot 10^{-5}$	Equation (4)	–
marginal utility of traveling (β_{trav}) [util/h]	-0.12	-0.0	-0.12	-0.40	-0.12
marginal utility of performing (β_{dur}) [util/h]			0.19		

marginal disutility caused by the martinal utility of time as a resource, plus a mode-specific 'additional' (direct) marginal disutility. The resulting mode-specific direct marginal utilities of traveling for MATSim scoring function are shown in Table 4.

Further, the ASCs for different modes are calibrated to capture the influence of variables not explicitly included in the scoring function. Along with this, to include the physical effort in bicycle and walk mode, the marginal utilities of distance for bicycle and walk, $\beta_{d,bicycle}$ and $\beta_{d,walk}$, are also calibrated.

In absence of any relevant data, the utility parameters of bicycle, car, and motorbike from urban and external traffic are assumed to be the same. For trucks, a different behavioral model is required, which is out of the scope of this study. However, for the scenario completion and to include the congestion effects from commercial vehicles, trucks are also included in the simulation with default utility parameters.[5] This means that they will search for their own fastest route and thus contribute to congestion, but they have no other choice dimension besides route, and will not be included into the economic analysis later.

3.2.3. Simulation setup

The modal splits of the urban travelers from reference study and initial plans are shown in Table 6. In order to replicate this modal split, mode choice is allowed for urban travelers and the ASCs are calibrated. The calibration is performed over 200 iterations together with CaDyTS in order to generate the synthetic plans for the external demand (see Section 2.2) and find destinations for commuters. For the calibration process, the maximum limit of plans in the choice set of an agent is set to 10. After calibrating with CaDyTS, only the best plans for each agent and in consequence only the destinations best matching the traffic counts are kept. The simulation is then continued for another 1000 iterations (i.e. overall 1200 iterations) to stabilize the urban and external demand in absence of CaDyTS.

Different so-called innovative modules are used for different sub-populations (urban and external).

(i) Urban: In a given iteration, 15% of the urban travelers are allowed to change their route, 10% are allowed to change mode and 5% are allowed to mutate the departure time of the activity. The mutation of the departure time of the activity is performed randomly between -2 to $+2$ h. The time mutation is turned off after CaDyTS calibration, i.e. the departure times of the urban travelers are then fixed.

(ii) External: In a given iteration, 15% of the agents from external traffic are allowed to change routes until innovation is turned off. After 200 iterations, the origin-destination pairs of the external demand are fixed.

---

[5] By default, the marginal utility of traveling, ASC, monetary distance rates for a mode are set to 0. This means, during a trip by mode truck, the agent will lose only opportunity cost of time ($= \beta_{dur} \cdot t_{trav}$).

Table 5. Calibrated utility parameters.

parameter	bicycle	car	motorbike	PT	walk
ASC (util)	0.0	−0.6	−0.58	−0.545	0.0
$\beta_{d,mode}$ (util/m)	−0.00011	−	−	−	−0.00012

Innovation is used until 80% of iteration (i.e., initially for iterations 1 to 160, and then for iterations 201 to 1000). The remaining agents until 80% of the iterations and all agents afterwards chose a plan from their generated choice sets. This plan selection follows a probability distribution which converges to a multinomial logit model (Nagel and Flötteröd 2012).

4. Calibration results

In this section, the results of the calibration are presented and the modal splits from reference study, initial plans and calibrated demand are compared. Afterwards, the real-world traffic counts are compared with the simulation counts. In order to understand the impact of the income-dependent scoring function, a comparison of income-dependent distance distribution from first and last iterations are presented.
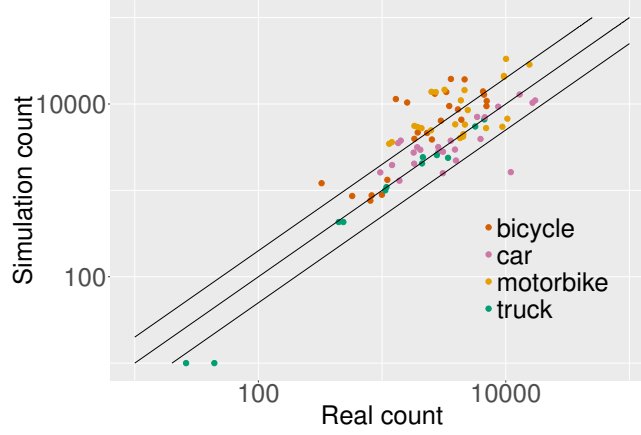
4.1. Calibrated utility parameters

$$S_{trav,bicycle} = -0.00 - \frac{0.12}{h} \cdot t_{trav} - \frac{0.00011}{m} \cdot d_{trav}$$

$$S_{trav,car} = -0.60 - \frac{0.0}{h} \cdot t_{trav} - \frac{3.7 \cdot 10^{-5}}{m} \cdot \frac{\bar{y}}{y_j} \cdot d_{trav}$$

$$S_{trav,motorbike} = -0.58 - \frac{0.12}{h} \cdot t_{trav} - \frac{1.6 \cdot 10^{-5}}{m} \cdot \frac{\bar{y}}{y_j} \cdot d_{trav} \qquad (8)$$

$$S_{trav,PT} = -0.545 - \frac{0.40}{h} \cdot t_{trav} - \gamma_{d,PT} \cdot \frac{\bar{y}}{y_j} \cdot d_{trav}$$

$$S_{trav,walk} = -0.00 - \frac{0.12}{h} \cdot t_{trav} - \frac{0.00012}{m} \cdot d_{trav}$$

The (manually) calibrated ASCs for all modes and marginal utility of distance for bicycle and walk modes are shown in Table 5 and Equation (8). The value of $\gamma_{d,PT}$ in Equation (8) is given by Equation (4). The ASCs for bicycle and walk modes are estimated to zero, which can be interpreted as no initial impedance. Car/motorbike and PT often have some initial overhead either in terms of getting the car out of the garage or in terms of walking to a PT stop. In this scenario, walking to PT stop is marginally less burdensome as getting the car/motorbike out of the garage/parking location. As a consequence of mode choice, the share of walk mode increases (see Table 6), which can be controlled either by a negative ASC or by having marginal utility of distance for walk mode ($\beta_{d,walk}$). The former has less significance for the walk mode and therefore the latter is chosen. In contrast to bicycle, the walk mode is teleported and thus the utility for a person with walk mode is not affected by congestion. The marginal utility of distance for the walk mode ($\beta_{d,walk} = -1.2 \cdot 10^{-4}$ util/m) is estimated marginally higher than the marginal utility of distance for the bicycle mode ($\beta_{d,walk} = -1.1 \cdot$

Table 6. Modal splits for urban demand.

mode	reference study (TRIPP, iTrans, and VKS 2009)	initial urban plans from travel diaries; it.0	after calibration it.1200
bicycle	33%	29.0%	32.3%
car	2%	4.0%	2.7%
motorbike	14%	20.3%	14.7%
PT	22%	26.6%	21.7%
walk	29%	20.1%	28.6%

Figure 2. Comparison of 24 h simulation and real traffic counts.

10^{-4} util/m). This means, for walking 1 km, an agent will loose 0.12 util. At a speed of 5 km/h, it will take 12 min which could be used for performing an activity. Thus, the agent will loose 0.024 util $(= \beta_{trav,walk} \cdot 0.2 \text{ h})$ for walking and 0.038 util $(= \beta_{dur} \cdot 0.2 \text{ h})$ opportunity cost of time which could be used for performing an activity.

4.2. *Modal split*

A comparison of the modal splits at different stages is shown in Table 6. It can be observed that the modal share for the walk mode is significantly different in the reference study and in the initial plans. The aim of the calibration is to replicate the modal shares from the reference study. Clearly, the modal split after calibration (column 'it.1200' in Table 6) has close resemblance with the reference study.

4.3. *Traffic counts*

Figure 2 shows the comparison of average weekday real counts and average weekday simulation counts after 1200 iterations. In the first step, CaDyTS pushes agents on the routes by adding a correction factor (Equation (2)) to the scoring function such that the simulation counts match the measured counts. Afterwards, in absence of the CaDyTS correction factor, the simulation counts for motorbike and bicycle become higher than the real counts and simulation counts for car and truck have a good match with real counts (see Figure 2). Eventually, the calibration results after 1200 iterations provide a good fit for modal split and synthetic plans for external traffic.

12

4.4. *Income-dependent distance distribution*
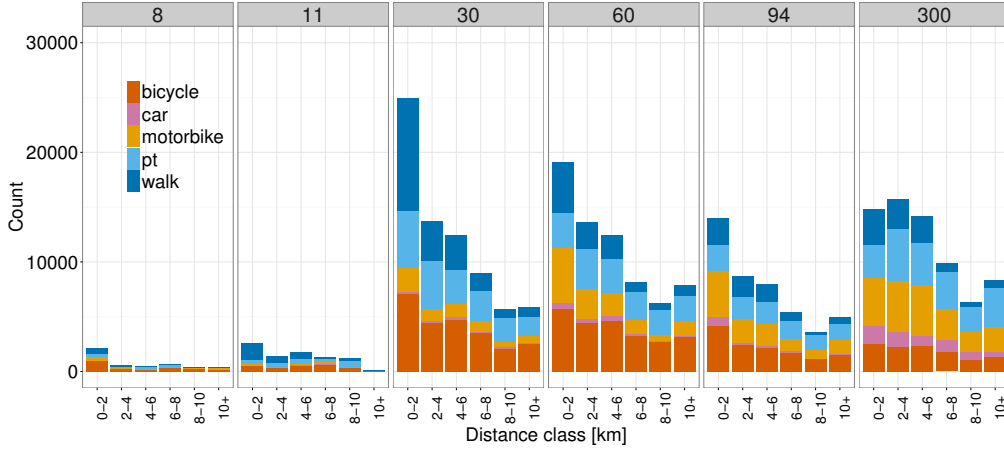
(a) it.0 (initial plans)
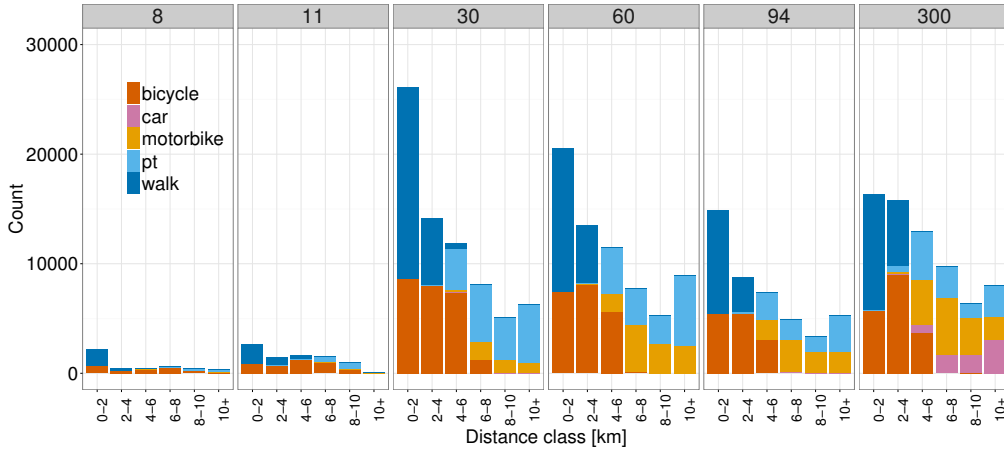
(b) it.1200 (calibrated plans)

Figure 3. Income-dependent distance distributions for initial plans and calibrated plans. The x- and y-axes depict the distance classes (in km) and number of trips respectively. The average income (in USD/month) is shown at the top of each frame.

In order to understand the impact of the income-dependent scoring function for different modes, the income-distance distribution is plotted in Figure 3. The income attributes are taken from the initial trip diaries and trip distances are the direct distances between origin and destination activities. The following observations are made:

a) After the calibration, the car is restricted to high income groups. In contrast to the initial plans, now the car is used for the longer distances.

b) PT is used mainly for longer distances (> 4 km), whereas bicycle and walk modes are used for relatively shorter distances (< 6 km). A few longer bicycle trips can also be observed for households with a very low income.

c) To replicate the modal share from the reference study, the scenario is calibrated such that the share of walk trips is about 8% higher after the calibration (see Table 6). A higher share of walk trips (relatively shorter distance i.e., < 4 km) can be noticed in the Figure 3(b). Additionally, the scoring function forces the

13

impractical longer (> 8 km) walk trips to more plausible modes. A similar effect is also observed for the longer bicycle trips from higher income groups.

Overall one can observe that several irregularities from the travel diaries are fixed in the calibrated plans which is suitable for policy testing.

5.   Conclusions

This study addresses the difficulties in the model development and validation due to limited availability of the data. The overall objectives of the study were to estimate the alternative specific constants (ASCs) in order to replicate the modal split in the reference study and include the perception of income levels in the utility function. In this direction, this study extended an approach to generate full day activity plans in heterogeneous traffic conditions. To simulate travel demand, an agent-based travel simulator was used, while for calibration, a Bayesian framework based calibration technique was used. A real-world scenario of Patna was used for this purpose. Diverse income levels were included in the utility function to filter out the errors in the survey data and to understand the impact of income levels on the decisions of travelers. In this approach, location choice was implicitly implemented to identify the initially unknown destinations based on the land use pattern. The calibrated ASCs show plausible values. With the help of income-based distance distributions, it was shown that the calibrated plans are feasible plans and free from the errors originated from the survey. In future, the authors wish to replace the manual calibration with an automatic calibration process using some optimization techniques (Agarwal, Flötteröd, and Nagel 2017).

References

Agarwal, A. 2012. "Agent based simulation of the travel demand for Patna City, India." Master's thesis, Indian Institute of Technology, Delhi, India.

Agarwal, A. 2017. "Mitigating negative transport externalities in industrialized and industrializing countries." PhD diss., TU Berlin, Berlin.

Agarwal, A., G. Flötteröd, and K. Nagel. 2017. "Calibration of behavioural parameters using optimization technique in an agent-based transport simulation." In *accepted for 6th Symposium of the European Association for Research in Transportation*, .

Agarwal, A., and G. Lämmel. 2016. "Modeling seepage behavior of smaller vehicles in mixed traffic conditions using an agent based simulation." *Transp. in Dev. Econ.* 2 (2): 1–12.

Agarwal, A., G. Lämmel, and K. Nagel. 2016. "Modelling of Backward Travelling Holes in Mixed Traffic Conditions." In *Traffic and Granular Flow '15*, edited by Victor L. Knoop and Winnie Daamen, 1st ed., Chap. 53, 419–426. Delft, NL: Springer International Publishing.

Agarwal, A., G. Lämmel, and K. Nagel. 2017. "Incorporating within link dynamics in an agent-

based computationally faster and scalable queue model." *Transportmetrica A: Transport Science* .

Agarwal, A., M. Zilske, K.R. Rao, and K. Nagel. 2015. "An elegant and computationally efficient approach for heterogeneous traffic modelling using agent based simulation." *Procedia Computer Science* 52 (C): 962–967.

Balmer, M., K. Meister, M. Rieser, K. Nagel, and K.W. Axhausen. 2008. "Agent-based simulation of travel demand: Structure and computational performance of MATSim-T." In *Innovations in Travel Modeling (ITM) '08*, Portland, OR, Jun. Also VSP WP 08-07, see http://www.vsp.tu-berlin.de/publications.

Barmpounakis, E. N., E. I. Vlahogianni, J. C. Golias, and A. Babinec. 2017. "How accurate are small drones for measuring microscopic traffic parameters?" *Transportation Letters* 1–9.

Bell, M. G. H. 1983. "The Estimation of an Origin-Destination Matrix from Traffic Counts." *Transportation Science* 17 (2): 198–217.

Bliemer, M. C. J. 2007. "Dynamic Queuing and Spillback in Analytical Multiclass Dynamic Network Loading Model." *Transportation Research Record: Journal of the Transportation Research Board* 2029: 14–21.

Cascetta, E., D. Inaudi, and G. Marquis. 1993. "Dynamic estimators of origin-destination matrices using traffic counts." *Transportation Science* 27 (4): 363–373.

Cetin, N., A. Burri, and K. Nagel. 2003. "A Large-Scale Agent-Based Traffic Microsimulation Based On Queue Model." In *Swiss Transport Research Conference (STRC)*, Monte Verita, Switzerland. See http://www.strc.ch, http://www.strc.ch.

Chandra, S., and P. K. Sikdar. 2000. "Factors affecting PCU in mixed traffic situations on urban roads." *Road and transport research* 9 (3): 40–50.

Charypar, D., and K. Nagel. 2005. "Generating complete all-day activity plans with genetic algorithms." *Transportation* 32 (4): 369–397.

Chen, Huey-Kuo, and Che-Fu Hsueh. 1998. "A model and an algorithm for the dynamic user-optimal route choice problem." *Transportation Research Part B: Methodological* 32 (3): 219–234.

Chen, Jingmin, and Michel Bierlaire. 2014. "Probabilistic Multimodal Map Matching With Rich Smartphone Data." *Journal of Intelligent Transportation Systems* 19 (2): 134–148.

Chung, Eui-Hwan, and Amer Shalaby. 2005. "A trip reconstruction tool for GPS-based personal travel surveys." *Transportation Planning and Technology* 28 (5): 381–401.

Currin, Thomas R. 2012. *Introduction to traffic engineering: a manual for data collection and analysis*. Cengage Learning.

Domencich, T., and D. L. McFadden. 1996. *Urban travel demand: a behavioral analysis*. North-Holland Publishing Company. https://eml.berkeley.edu/~mcfadden/travel.html.

Flötteröd, G. 2010. *Cadyts – Calibration of dynamic traffic simulations – Version 1.1.0 manual*. Transport and Mobility Laboratory, École Polytechnique Fédérale de Lausanne. http://home.abe.kth.se/~gunnarfl/files/cadyts/Cadyts_manual_1-1-0.pdf.

Flötteröd, G., M. Bierlaire, and K. Nagel. 2011. "Bayesian demand calibration for dynamic traffic simulations." *Transportation Science* 45 (4): 541–561.

Flötteröd, G., Y. Chen, and K. Nagel. 2011. "Behavioral Calibration and Analysis of a Large-Scale Travel Microsimulation." *Networks and Spatial Economics* 12 (4): 481–502.

Franklin, J.P. 2006. "The distributional effects of transportation policies: The case of a bridge toll for Seattle." PhD diss., University of Washington, Seattle.

Gawron, C. 1998. "An Iterative Algorithm to Determine the Dynamic User Equilibrium in a Traffic Simulation Model." *International Journal of Modern Physics C* 9 (3): 393–407.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *The Public Opinion Quarterly* 70 (5): 646–675. http://www.jstor.org/stable/4124220.

Hood, J., E. Sall, and B. Charlton. 2011. "A GPS-based bicycle route choice model for San Francisco, California." *Transportation Letters* (3): 63–75.

Horni, A., K. Nagel, and K. W. Axhausen, eds. 2016. *The Multi-Agent Transport Simulation MATSim*. Ubiquity, London. http://matsim.org/the-book.

Iqbal, Shahadat, Charisma Choudhury, Pu Wang, and Marta C González. 2014. "Development

of origin-destination matrices using mobile phone call data." *Transportation Research Part C* 40: 63–74.

IRC:SP:30. 2009. *Manual on economic evaluation of highway projects in India*. New Delhi, India: Indian Roads Congress.

Kickhöfer, B., and K. Nagel. 2016. "Microeconomic Interpretation of MATSim for Benefit-Cost Analysis." In *The Multi-Agent Transport Simulation MATSim*, edited by A. Horni, K. Nagel, and K. W. Axhausen, Chap. 51. Ubiquity, London. http://matsim.org/the-book.

Kumar, C. V. P., D. Baus, and B. Maitra. 2004. "Modeling generalized cost of travel for rural bus users: a case study." *Journal of Public Transportation* 7 (2): 59–72.

Kuwahara, Masao, and Edward C. Sullivan. 1987. "Estimating origin-destination matrices from roadside survey data." *Transportation Research Part B: Methodological* 21 (3): 233–248.

Lam, William H.K., and Hai-Jun Huang. 1995. "Dynamic user optimal traffic assignment model for many to one travel demand." *Transportation Research Part B: Methodological* 29 (4): 243–259.

Lee, Richard J., Ipek N. Sener, and James A. Mullins III. 2016. "An evaluation of emerging data collection technologies for travel demand modeling: from research to practice." *Transportation Letters* 8 (4): 181–193.

Moyo Oliveros, M., and K. Nagel. 2012. *Automatic Calibration of Microscopic, Activity-Based Demand for a Public Transit Line*. Annual Meeting Preprint 12-3279. Washington, D.C.: Transportation Research Board. Also VSP WP 11-13, see http://www.vsp.tu-berlin.de/publications.

Nagel, K., and G. Flötteröd. 2012. "Agent-based traffic assignment: Going from trips to behavioural travelers." In *Travel Behaviour Research in an Evolving World – Selected papers from the 12th international conference on travel behaviour research*, edited by R.M. Pendyala and C.R. Bhat, 261–294. International Association for Travel Behaviour Research.

Rieser-Schüssler, Nadine. 2012. "Capitalising modern data sources for observing and modelling transport behaviour." *Transportation Letters* 4 (2): 115–128.

Shen, Li, and Peter R. Stopher. 2014. "Review of GPS Travel Survey and GPS Data-Processing Methods." *Transport Reviews* 34 (3): 316–334.

Simon, P.M., J. Esser, and K. Nagel. 1999. "Simple queueing model applied to the city of Portland." *International Journal of Modern Physics* 10 (5): 941–960.

Stipancic, J., L. MIranda-Moreno, A. Labbe, and N. Saunier. 2017. "Measuring and visualizing space–time congestion patterns in an urban road network using large-scale smartphone-collected GPS data." *Transportation Letters* 1–11.

Szeto, W. 2008. "Enhanced lagged cell-transmission model for dynamic traffic assignment." *Transportation Research Record: Journal of the Transportation Research Board* 2085: 76–85.

Szeto, W., and S. Wong. 2012. "Dynamic traffic assignment: model classifications and recent advances in travel choice principles." *Open Engineering* 2 (1): 1–18.

Szeto, W.Y., Y. Jiang, and A. Sumalee. 2011. "A cell-based model for multi-class doubly stochastic dynamic traffic assignment." *Computer-Aided Civil and Infrastructure Engineering* 26 (8): 595–611.

TRIPP, iTrans, and VKS. 2009. *Comprehensive mobility plan for Patna urban agglomeration area*. Technical Report. Department of Urban Development. Government of Bihar.

van Zuylen, H., and L.G. Willumsen. 1980. "The most likely trip matrix estimated from traffic counts." *Transportation Research* 14B: 281–293.

Wolf, J. 2000. "Using GPS data loggers to replace travel diaries in the collection of travel data." PhD diss., Georgia Institute of Technology.

Ziemke, D., and K. Nagel. 2017. *Development of a fully synthetic and open scenario for agent-based transport simulations – The MATSim Open Berlin Scenario*. VSP Working Paper 17-12. TU Berlin, Transport Systems Planning and Transport Telematics. URL http://www.vsp.tu-berlin.de/publications.

Ziemke, D., K. Nagel, and C. Bhat. 2015. "Integrating CEMDAP and MATSim to increase the transferability of transport demand models." *Transportation Research Record* 2493: 117–125.

Zilske, M., and K. Nagel. 2015. "A Simulation-based Approach for Constructing All-day Travel

Chains from Mobile Phone Data." *Procedia Computer Science* 52: 468–475.

Zimowski, M., R. Tourangeau, R. Ghadialy, and S. Pedlow. 1997. *Nonresponse in household travel surveys.* Technical Report. Federal Highway Administration.

Appendix A. Patna external demand

The external demand for Patna scenario is generated as follows.

Table A1. An example of hourly classified traffic counts data.

time bin	car	motorbike	truck	bicycle	total
1	34	5	142	1	182
...	...	...	...	...	...
6	43	38	210	68	359
7	48	93	139	101	381
8	76	123	141	137	477
9	56	33	42	36	167
...	...	...	...	...	...
22	115	55	165	10	345
23	95	40	225	3	363
24	49	16	186	1	252

1) TRIPP, iTrans, and VKS (2009) provide hourly classified traffic counts data for all counting stations in both (inbound and outbound) directions (see Table A1 for an example). For each mode, the daily sum of hourly inbound and outbound counts must be equal, if this is not the case, the counts are adjusted. For instance, total inbound car count is 990 and outbound count is 1000, thus, the outbound counts are reduced by a factor calculated as $(1000 - 990)/990$.

Table A2. Share of through and commuters traffic.

Outer cordon location	Share of ...	
	commuters traffic	through traffic
OC1	0.70	0.30
OC2	0.58	0.42
OC3	0.94	0.06
OC4	0.66	0.34
OC5	0.76	0.24
OC6	0.86	0.14
OC7	0.95	0.05

2) Further, the directional split for each counting station is available (see Table A2). In absence of the classified hourly factors, the directional split is used together with the adjusted hourly classified counts (from step 1) to get the hourly modal counts for commuters and through traffic. E.g., at OC1, for time bin 2, the car count is 100; 70% of this will be commuters and the remaining 30 will be through traffic.

3) Further, Patna CMP also provides an origin-destination (OD) matrix for through traffic which helps to determine the origin and destination of the through trip. Again, in absence of the hourly classified OD matrix, the through traffic counts obtained in step 2 are used along with the OD matrix (see Table A3) to get the through trips. From the example in step 2, of the 30 through car trips that originate at OC1 in time bin 2, 49% trips ($\approx$ 15) terminate at OC4, 15% trips ($\approx$ 5) terminate at OC5, etc.

Table A3. Origin-destination (O-D) matrix for through traffic.

O-D	OC1	OC2	OC3	OC4	OC5	OC6	OC7
OC1	0%	0%	2%	49%	15%	3%	31%
OC2	1%	0%	0%	84%	5%	0%	10%
OC3	19%	4%	0%	4%	17%	23%	33%
OC4	76%	16%	0%	0%	3%	0%	5%
OC5	35%	7%	4%	38%	0%	8%	8%
OC6	30%	7%	23%	0%	13%	0%	27%
OC7	34%	7%	0%	9%	50%	0%	0%