Investigating Micro-simulation Error in Activity-based Travel Demand Forecasting using Confidence Intervals

Qiong Bao, Bruno Kochan, Tom Bellemans, Davy Janssens, and Geert Wets*

Hasselt University
Transportation Research Institute (IMOB)
Wetenschapspark 5 bus 6
3590 Diepenbeek
Belgium

E-mails:
{qiong.bao, bruno.kochan, tom.bellemans, davy.janssens, geert.wets}@uhasselt.be

*Corresponding author: Geert Wets
Tel: +32(0)11269158
Fax: +32(0)11269198

Word count: 4904+9*250=7154
Number of Figures: 3
Number of Tables: 6
Submission date: 13/11/2012

Abstract

Activity-based models of travel demand using micro-simulation approach inevitably include stochastic error that is caused by the statistical distributions of random components. As a result, running a traffic micro-simulation model several times with the same inputs will obtain different outputs. In order to take the variation of outputs in each model run into account, a common approach is to run the model multiple times and to use the average value of the results. The question then becomes: what is the minimum number of model runs required to reach a stable result (i.e., with a certain level of confidence that the obtained average value can only vary within an acceptable interval). In this study, systematic experiments are carried out by using the FEATHERS framework, an agent-based micro-simulation model particularly developed for Flanders, Belgium. Six levels of geographic detail are taken into account, which are Building block level, Subzone level, Zone level, Superzone level, Province level, and the whole Flanders. Three travel indices, i.e., the average daily number of trips per person, the average daily distance travelled per person, and the average daily number of activities per person, as well as their corresponding segmentations, are estimated by running the model 100 times. The results show that the more detailed geographical level is considered, the larger the number of model runs is needed to ensure confidence of a certain percentile of zones at this level to be stable. In addition, based on the time-dependent origin-destination table derived from the model output, traffic assignment is performed by loading it onto the Flemish road network, and the total vehicle kilometres travelled in the whole Flanders are computed subsequently. The stable results at the Flanders level provides model users with confidence that application of the FEATHERS at an aggregated level only requires limited model runs.

1 INTRODUCTION

Travel demand modelling was first developed in the late 1950s as a means to do highway planning. The four-step model, as the exemplification of the conventional trip-based approach, is the primary tool for forecasting future demand and performance of regional transportation systems (*1*). However, traditional trip-based approaches consider the trip as the unit of analysis, and the trip chains made by an individual are treated as separate, independent entities in the analysis, which often leads up to failure of recognizing the existence of linkages among trips. In some instances, the forecasts of trip-based approaches have proved to be inaccurate due to such an inappropriate representation of travel behaviour relationships (*2*). In the 1970s, the activity-based approach emerged, which explicitly recognizes and addresses the inability of conventional trip-based approach to reflect underlying human behaviour in general, and travel behaviour in particular. The approach is a richer, more holistic framework in which travel is analyzed as daily or multi-day patterns of behaviour related to and derived from differences in lifestyles and activity participation among the population (*2*). A full activity-based model of travel demand predicts which activities (activity participation) are conducted where (destination choice), when (timing), for how long (duration), which chain of transport modes is involved (mode choice), travel party (travel arrangements and joint activity participation) and which route is chosen (route choice), subject to personal, household, spatial, temporal, institutional and space-time constraints (*3*). In the following 1990s, a rapid growth of interest in activity-based analysis has led to the development of several practical models, including TRAMSIMS (*4*), RAMBLAS (*5*), CEMDAP (*6*), FAMOS (*7*), ALBATROSS (*8-10*), and FEATHERS (*11*). The main contribution of these activity-based models is to offer an alternative to the four-step models of travel demand, better focusing on the consistency of the sub-models and proving increased sensitivity to a wider range of policy issues (*12*).

However, the activity-based models, focusing on activity-travel generation and activity scheduling decisions, use in most cases micro-simulation approach, in which heterogeneity and randomness are fundamental characteristics since they simulate individual activity patterns by drawing randomly from marginal and conditional probability distributions that are defined for the various choice facets that make up an activity pattern (*13*). As a result, running a traffic micro-simulation model several times with the same inputs will obtain different outputs due to the random number seed used in each run. In order to address practitioners' concerns about this variation, it is natural to run the traffic micro-simulation model multiple times, estimate the effects of stochastic error by analysing the variation of the outputs between the runs, and use the average value of these outputs for further analysis. The question then becomes: what is the minimum number of runs required to reach a stable result (i.e., with a certain level of confidence that the obtained average value can only vary within an acceptable interval)? In this respect, Castiglione et al. (*14*) investigated the extent of random variability in the San Francisco model (a micro-simulation model system used in actual planning applications since 2000) by running the model 100 times at three levels of geographic detail (i.e., zone level, neighborhood level, and county-wide level). The analysis was then conducted by showing how quickly the mean values of output variables such as the number of trips per person converge towards the final mean value (after 100 runs) as the number of simulation runs increases. However, only two zones and neighborhoods were considered in that study, which to a large extent limits the generalization of the conclusions drawn in that paper. In this study, we focus on the same issue but look for the answer one step further, which is to find the minimum number of model runs needed to enable at least a certain percentile of zones at different levels of geographic detail to reach a stable result.

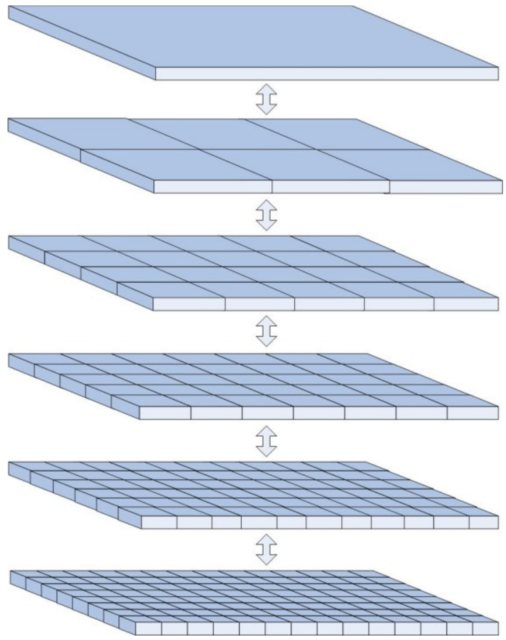
Systematic experiments are carried out by using the FEATHERS framework, an agent-based micro-simulation model particularly developed for Flanders, Belgium. Six different geographical levels are taken into account, which are Building block level, Subzone level, Zone level, Superzone level, Province level, and the whole Flanders level (see Section 2). By running the model 100 times, three travel indices, i.e., the average daily number of trips per person, the average daily distance travelled per person and the average daily number of activities per person, as well as their corresponding segmentations with respect to socio-demographic variables, transport mode choice, and activity types, are estimated. The results show that the more disaggregated geographical level is considered, the greater the number of model runs is needed to ensure confidence of a certain percentile of zones at this level to be stable. In addition, based on the time-dependent origin-destination table derived from the model output, traffic assignment is performed by loading it onto the Flemish road network, and the total vehicle kilometres travelled in the whole Flanders are computed subsequently. The relative stability of the results provide model users with confidence that application of the FEATHERS at an aggregated level only requires limited model runs.

The remaining of this paper is structured as follows. In Section 2, we briefly introduce the FEATHERS framework and the levels of geographic detail of Flanders, followed by the detailed elaboration of the experiment execution in Section 3. In Section 4, the analysis results are presented and further discussed. The paper ends with conclusions in Section 5.

2 FEATHERS FRAMEWORK FOR FLANDERS

FEATHERS (The Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS) (*11*) is a rule- and agent-based micro-simulation framework, particularly developed to facilitate the development of dynamic activity-based models for transport demand forecast in Flanders, Belgium. In this framework, a sequence of 26 decision trees, derived by means of the chi-squared automatic interaction detector (CHAID) algorithm, is used in the scheduling process and decisions are based on a number of attributes of the individual (e.g. age, gender), of the household (e.g. number of cars), and of the geographical zone (e.g. population density, number of shops) (*9*). For each agent (i.e., person) with its specific attributes, the model simulates whether an activity (e.g. shopping, working, leisure activity, etc.) is going to be carried out or not. Subsequently, the location, transport mode and duration of the activity are determined, taking into account the attributes of the individual. Based on the estimated schedules or activity travel patterns, travel demand can then be extracted and assigned to the transportation network. Currently, the FEATHERS framework is fully operational at six levels of geographic detail of Flanders, i.e., Building block (BB) level, Subzone level, Zone level, Superzone level, Province level, and the whole Flanders level, which are illustrated in Fig. 1.

In recent years, a number of applications have been carried out using the FEATHERS framework (see e.g., (*15-17*)). However, like other activity-based models, the FEATHERS framework is based on micro-simulation approach. Stochastic error thereby inherently exists, which requires systematic investigation with the purpose of better understanding the variability of simulation results and of the further development of this modelling framework.



Geographical Level	Number of zones	Average area (km^2)
Flanders	1	13709.24
Province	6	2284.87
Superzone	327	41.93
Zone	1145	11.97
Subzone	2386	5.75
BB	10521	1.30

FIGURE 1 Six levels of geographic detail of Flanders used in the FEATHERS.

3 METHODOLOGY

In this study, to estimate the impact of stochastic error of the FEATHERS framework at all the six levels of geographic detail of Flanders, 100 successive model runs are performed based on the 10% fraction of the full population. Computational time is thus to a great extent saved, but it still takes around 18 hours for a single model run at the BB level, the most disaggregated geographical scale.

After each model run, the prediction file, containing the whole activity travel pattern or schedule information for each agent, is generated, based on which the three travel indices (i.e., the average daily number of trips per person, the average daily distance travelled per person and the average daily number of activities per person) can be computed. Moreover, segmentations of these travel indices based on socio-demographic variables, transport mode choice, as well as activity types can be obtained.

Recall the main objective of this study, which is to determine the minimum number of model runs needed to ensure a certain percentile of zones at different geographical levels to reach a stable result (i.e., with a certain level of confidence that the obtained average value of each of these zones can only vary within an acceptable interval) concerning the travel indices. The concept of confidence interval (CI) is therefore adopted in this study, and the following equation is applied (*18*):

$$CI_{(1-\alpha)\%} = 2 \times t_{(1-\alpha/2),N-1} \frac{s}{\sqrt{N}} \quad (1)$$

where $CI_{(1-\alpha)\%}$ represents $(1-\alpha)\%$ confidence interval for the true average value; α is the probability of the true average value not lying within the confidence interval; $t_{(1-\alpha/2),N-1}$ is the Student's *t*-statistic for the probability of a two-sided error summing to α with *N*-1 degrees of freedom; *N* is the required number of model runs; and *s* denotes the estimated standard deviation of the results.

For the experiment, a 95% level of confidence is selected and the desired confidence interval, which acts as the predefined stable condition, is set as a 10% fraction of the final average value (after 100 runs) of the index (*X*) under study, i.e., $CI \le 0.1 \times \bar{X}_{100}$, where $\bar{X}_{100} = \sum_{i=1}^{100} X \big/ 100$. Also, the standard deviation of the results among 100 runs is used as the estimation of *s*.

Now, by using Eq. (1), an iterative process is applied for each zone to estimate the required minimum number of model runs in terms of the corresponding index under study. In short, it is necessary to iterate until the estimated number of model runs *N* matches the number of repetitions assumed when looking up the student's t statistic. In this way, the minimum number of FEATHERS runs needed to ensure a certain percentile of zones at different geographical levels to achieve stable results with respect to the corresponding index can be derived.

Furthermore, by considering the socio-demographic variables gender (two categories: male and female) and age (five categories: 18-34 years, 35-54 years, 55-64 years, 65-74 years, and 75+ years) as well as four types of transport modes (i.e., car as driver, car as passenger, slow mode, and public transport) and four types of activities (i.e., home-related activity, work-related activity, shopping activity, and touring activity), the required minimum number of FEATHERS runs with respect to these segmentations can be obtained respectively.

In addition, based on the time-dependent origin-destination table derived from the model output, traffic assignment can be performed by loading it onto the Flemish road network, and the vehicle kilometres travelled at the whole Flanders can be studied subsequently.

4 RESULTS AND DISCUSSION

In this section, the results of the experiment on the average daily number of trips per person, the average daily distance travelled per person, and the average daily number of activities per person, as well as their related segmentations at all the geographical levels of Flanders are presented. The vehicle kilometres travelled on the Flemish road network after traffic assignment is provided subsequently.

4.1 Travel indices

According to Eq. (1), the required minimum number of FEATHERS runs for each zone at all the geographical levels can be calculated based on the predefined stable condition. Fig. 2 illustrates the minimum number of model runs needed to enable different percentiles of zones of each geographical level to reach the stability with respect to the average daily number of trips per person, the average daily distance travelled per person, and the average daily number of activities per person, respectively.

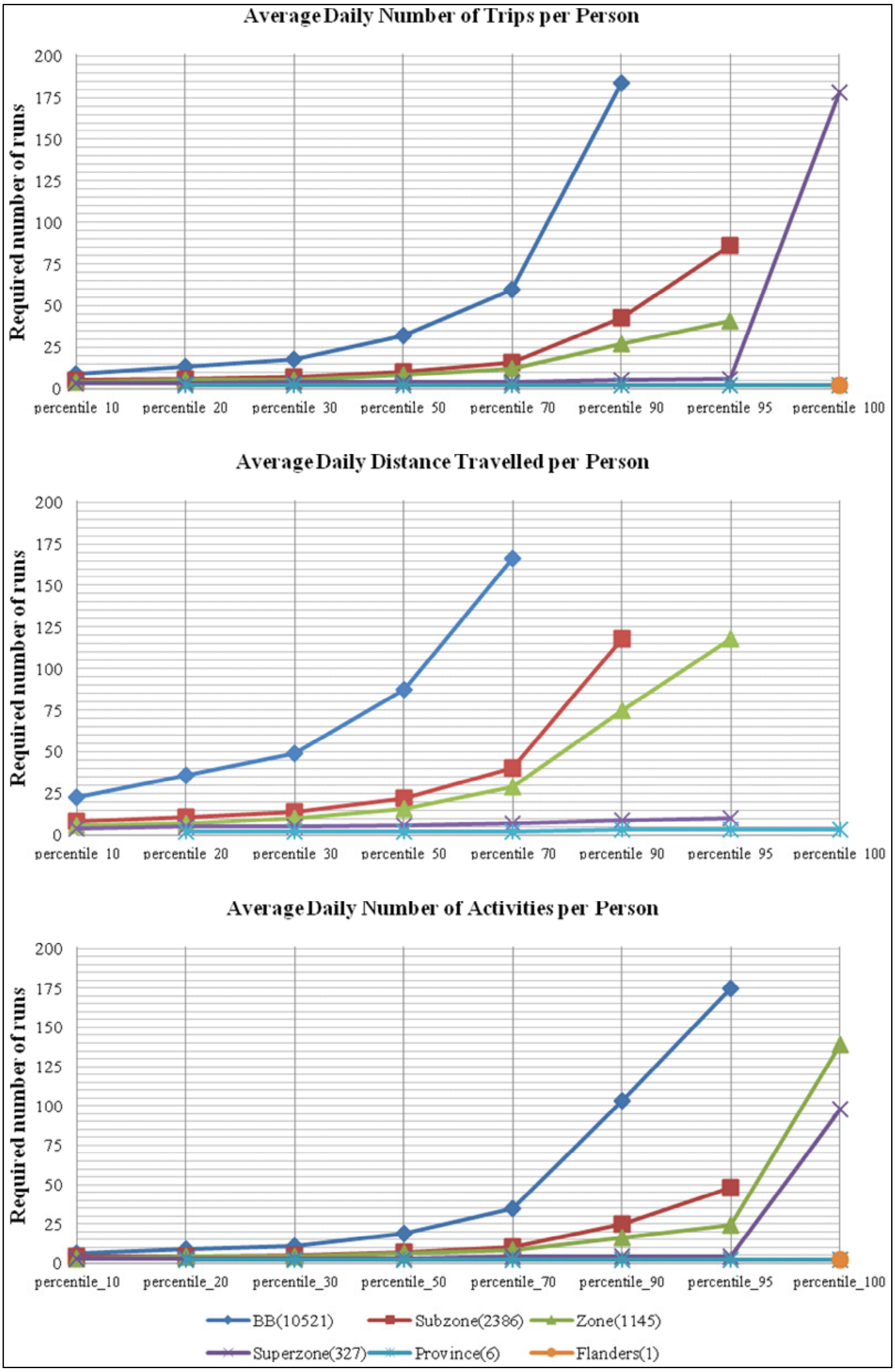


FIGURE 2 The required minimum number of model runs for different percentiles of stable zones at 6 geographical levels on three travel indices.

In general, the required minimum number of runs for the daily distance travelled is larger than that for the daily number of trips, which is in turn larger than that for the daily number of activities, especially for the lower geographical levels, such as the BB level, the Subzone level, and the Zone level. This can be mainly accounted for by the fact that in the FEATHERS framework, the type of activities is firstly scheduled, followed by the determination of activity locations. The stochastic error is therefore accumulated by executing each of the above procedures (*9,19*).

Moreover, for all the three indices, with a decrease in the geographical aggregation level, the required minimum number of model runs to enable the certain percentile of zones to achieve the predefined stable condition is increasing, which means that relative to a highly aggregated geographical level, it's more difficult for a lower level to make the same percentile of zones reach stability. In other words, with a certain number of model runs, a lower geographical level can only guarantee a smaller percentile of zones to reach stable status. Taking the daily number of trips as an example, at both the Flanders and the province levels, the sample mean of this index has negligible variation, thereby only limited number of runs is needed to ensure all the zones in these levels to be stable. When it comes to the Superzone level, also few runs are needed if only 95% of the zones are required to be stable. However, if the stability of all the zones at this level is the requirement, then around 180 model runs have to be performed. The situation becomes worse when lower geographical levels are taken into account. At the final BB level, 180 model runs can only ensure 90% of the zones to be stable, and within 100 runs, only 70% of the zones can be guaranteed in terms of their stability. It is therefore a dilemma between exploring in more complex or detailed circumstances and seeking for more reliable results. One compromising solution is to set another relatively achievable confidence interval condition for the zones with high variation, especially when these zones are not involved in the study area.

4.2 Segmentations

In order to illustrate the impact of segmentations of the population on the required number of model runs, the above travel indices are disaggregated based on socio-demographic variables (gender and age), transport mode choice, as well as different activity types. The results are presented in the following sections.

4.2.1 Gender

Fig. 3 illustrates the results of gender segmentation related to the average daily number of trips per person and the average daily distance travelled per person. As can be seen, the required minimum number of model runs for either male or female is a little bit larger than that of the overall travel indices for each percentile due to the classification by gender. Moreover, the female group needs a relatively larger number of runs for each percentile of zones to reach the predefined stability than the male group. It can be partly attributed to the fact that as a whole the female group generates a relatively smaller number of trips, and distance travelled as well than that of the male group.

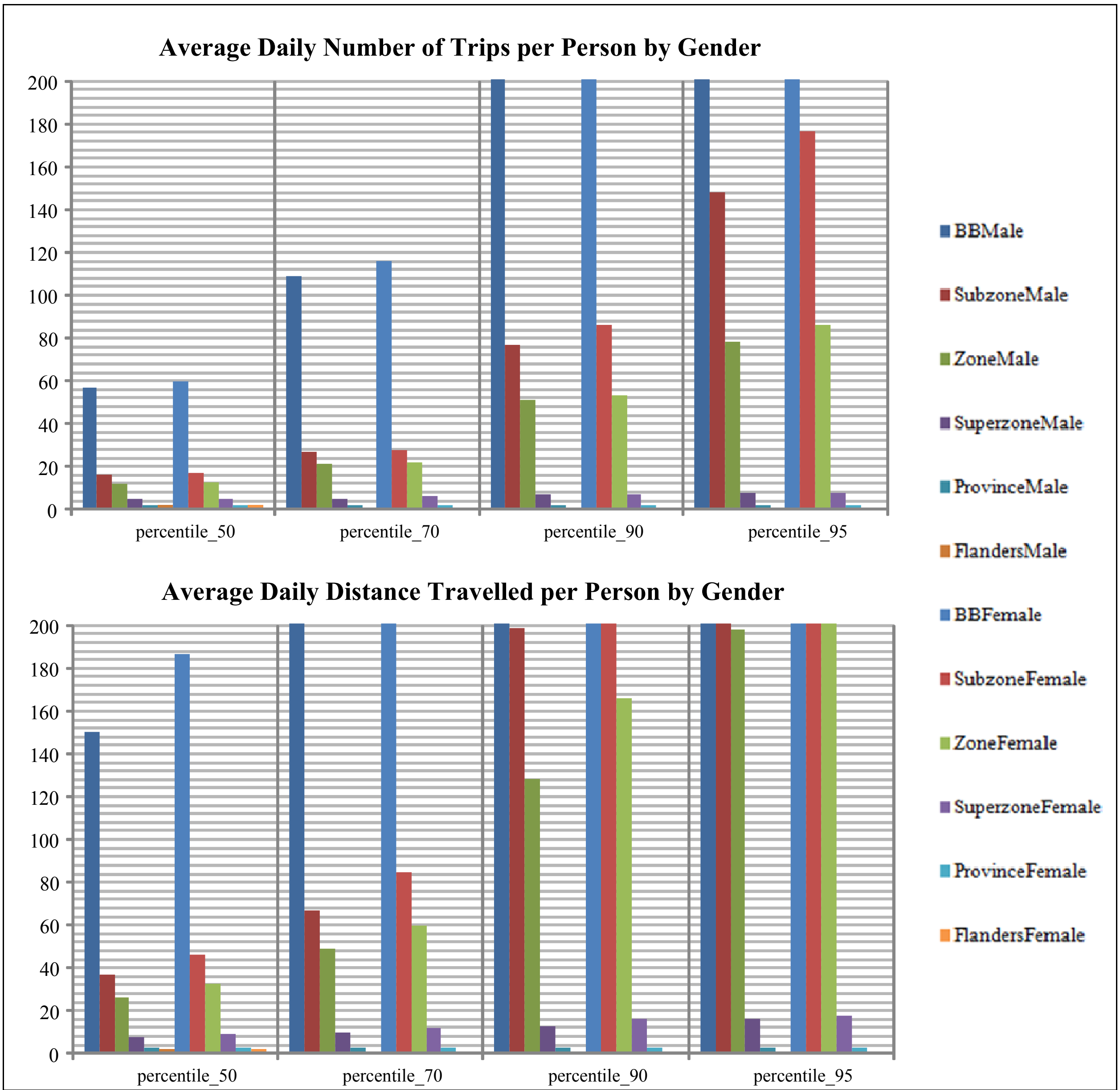


FIGURE 3 The required minimum number of model runs for different percentiles of stable zones at 6 geographical levels by gender on average daily number of trips and distance travelled per person.

4.2.2 Age

When age categories are considered with respect to the same travel indices analyzed in Section 4.2.1, the required minimum number of model runs for different percentiles is significantly increased, especially at the highly disaggregated geographical levels. Whereas at the Flanders and the Province levels, less than 5 runs are needed for both indices, even when the full percentile is under requirement (see Tables 1 and 2). Moreover, concerning the lower geographical levels, it is interesting to see that the required runs for the first two age categories (i.e., 18-34 years and 35-54 years) are apparently less than that for the following two age categories (i.e., 55-64 years and 65-74 years), which are further less than that of the last age category, i.e., over 75 years. This dissimilarity between different age groups can be explained by the fact that the first two age groups involve a larger population than the age groups 55-64 years and 65-74 years, which also involve a larger population than the eldest age group. Such a situation potentially increases the instability of the index under study with respect to the elder age group because less population normally implies a fewer number of trips and distance travelled.

TABLE 1 The required minimum number of model runs for different percentiles of stable zones at 6 geographical levels by age on average daily number of trips per person.

BB (10521)	Nr. of persons	required minimum nr of runs				**Subzone (2386)**	Nr. of persons	required minimum nr of runs			
		p_50	p_70	p_90	p_100			p_50	p_70	p_90	p_100
18-34	119657	81	161	>200	>200	18-34	119657	26	47	138	>200
35-54	181022	59	113	>200	>200	35-54	181022	17	31	84	>200
55-64	67781	143	>200	>200	>200	55-64	67781	53	94	>200	>200
65-74	63261	186	>200	>200	>200	65-74	63261	70	129	>200	>200
75+	47409	>200	>200	>200	>200	75+	47409	127	>200	>200	>200
Zone (1145)	Nr. of persons	required minimum nr of runs				**Superzone (327)**	Nr. of persons	required minimum nr of runs			
		p_50	p_70	p_90	p_100			p_50	p_70	p_90	p_100
18-34	119657	20	36	98	>200	18-34	119657	7	8	10	42
35-54	181022	13	23	61	>200	35-54	181022	5	6	8	>200
55-64	67781	41	77	>200	>200	55-64	67781	10	13	21	>200
65-74	63261	52	103	>200	>200	65-74	63261	13	17	26	>200
75+	47409	91	180	>200	>200	75+	47409	23	30	45	>200
Province (6)	Nr. of persons	required minimum nr of runs				**Flanders (1)**	Nr. of persons	required minimum nr of runs			
		p_50	p_70	p_90	p_100			p_50	p_70	p_90	p_100
18-34	119657	2	2	3	3	18-34	119657	--	--	--	2
35-54	181022	2	2	2	2	35-54	181022	--	--	--	2
55-64	67781	3	3	3	3	55-64	67781	--	--	--	2
65-74	63261	3	3	3	3	65-74	63261	--	--	--	2
75+	47409	3	3	3	3	75+	47409	--	--	--	3

Note: In this table, p_50 represents percentile 50. The same goes to p_70, p_90, and p_100.
At the Flanders level, there is only one geographical zone, therefore the concept of p_50, p_70 and p_90 is not applicable.

TABLE 2 The required minimum number of model runs for different percentiles of stable zones at 6 geographical levels by age on average daily distance travelled per person.

BB (10521)	Nr. of persons	required minimum nr of runs				**Subzone (2386)**	Nr. of persons	required minimum nr of runs			
		p_50	p_70	p_90	p_100			p_50	p_70	p_90	p_100
18-34	119657	>200	>200	>200	>200	18-34	119657	69	124	>200	>200
35-54	181022	158	>200	>200	>200	35-54	181022	41	77	>200	>200
55-64	67781	>200	>200	>200	>200	55-64	67781	156	>200	>200	>200
65-74	63261	>200	>200	>200	>200	65-74	63261	>200	>200	>200	>200
75+	47409	>200	>200	>200	>200	75+	47409	>200	>200	>200	>200
Zone (1145)	Nr. of persons	required minimum nr of runs				**Superzone (327)**	Nr. of persons	required minimum nr of runs			
		p_50	p_70	p_90	p_100			p_50	p_70	p_90	p_100
18-34	119657	50	97	>200	>200	18-34	119657	13	16	22	127
35-54	181022	29	56	156	>200	35-54	181022	9	10	15	>200

55-64	67781	113	>200	>200	>200	55-64	67781	25	34	54	>200
65-74	63261	166	>200	>200	>200	65-74	63261	38	49	76	>200
75+	47409	>200	>200	>200	>200	75+	47409	70	95	149	>200
Province (6)	Nr. of persons	required minimum nr of runs				**Flanders (1)**	Nr. of persons	required minimum nr of runs			
		p_50	p_70	p_90	p_100			p_50	p_70	p_90	p_100
18-34	119657	3	3	3	3	18-34	119657	--	--	--	2
35-54	181022	3	3	3	3	35-54	181022	--	--	--	2
55-64	67781	3	3	3	3	55-64	67781	--	--	--	2
65-74	63261	3	3	4	4	65-74	63261	--	--	--	3
75+	47409	4	4	4	5	75+	47409	--	--	--	3

Note: In this table, p_50 represents percentile 50. The same goes to p_70, p_90, and p_100.
At the Flanders level, there is only one geographical zone, therefore the concept of p_50, p_70 and p_90 is not applicable.

4.2.3 Transport modes

In addition to the socio-demographical variables, research on the mode split is also important from the practitioner's point of view. In this study, four different transport modes, i.e., car as driver, car as passenger, slow mode, and public transport are considered. The results are shown in Table 3 and Table 4. We find that the most frequently used transport mode in Flanders, i.e., the car as driver, needs the fewest number of model runs to reach the predefined stable condition for both the trip and the distance related indices at any geographical level and for any required percentile of zones. On the contrary, the public transport appears to be the mode with the highest variation since the largest number of model runs are needed to achieve the predefined confidence interval.

TABLE 3 The required minimum number of model runs for different percentiles of stable zones at 6 geographical levels by transport modes on average daily number of trips per person.

BB (10521)	required minimum nr of runs				**Subzone (2386)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Car as Driver	62	119	>200	>200	Car as Driver	16	29	87	>200
Car as Passenger	>200	>200	>200	>200	Car as Passenger	74	136	>200	>200
Slow Mode	184	>200	>200	>200	Slow Mode	47	88	>200	>200
Public Transport	>200	>200	>200	>200	Public Transport	161	>200	>200	>200
Zone (1145)	required minimum nr of runs				**Superzone (327)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Car as Driver	12	22	54	>200	Car as Driver	5	6	7	>200
Car as Passenger	50	99	>200	>200	Car as Passenger	13	17	24	>200
Slow Mode	35	68	174	>200	Slow Mode	9	12	17	>200
Public Transport	112	>200	>200	>200	Public Transport	25	33	50	>200
Province (6)	required minimum nr of runs				**Flanders (1)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Car as Driver	2	2	2	2	Car as Driver	--	--	--	2
Car as Passenger	3	3	3	3	Car as Passenger	--	--	--	2
Slow Mode	3	3	3	3	Slow Mode	--	--	--	2
Public Transport	3	3	3	3	Public Transport	--	--	--	3

Note: In this table, p_50 represents percentile 50. The same goes to p_70, p_90, and p_100.
At the Flanders level, there is only one geographical zone, therefore the concept of p_50, p_70 and p_90 is not applicable.

TABLE 4 The required minimum number of model runs for different percentiles of stable zones at 6 geographical levels by transport modes on average daily distance travelled per person.

BB (10521)	required minimum nr of runs				**Subzone (2386)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Car as Driver	139	>200	>200	>200	Car as Driver	33	61	188	>200
Car as Passenger	>200	>200	>200	>200	Car as Passenger	143	>200	>200	>200
Slow Mode	>200	>200	>200	>200	Slow Mode	>200	>200	>200	>200
Public Transport	>200	>200	>200	>200	Public Transport	>200	>200	>200	>200
Zone (1145)	required minimum nr of runs				**Superzone (327)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Car as Driver	24	44	119	>200	Car as Driver	7	9	12	>200
Car as Passenger	100	190	>200	>200	Car as Passenger	22	31	44	>200
Slow Mode	160	>200	>200	>200	Slow Mode	34	49	74	>200
Public Transport	173	>200	>200	>200	Public Transport	38	50	79	>200
Province (6)	required minimum nr of runs				**Flanders (1)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Car as Driver	3	3	3	3	Car as Driver	--	--	--	2
Car as Passenger	3	3	3	3	Car as Passenger	--	--	--	2
Slow Mode	3	3	3	4	Slow Mode	--	--	--	3
Public Transport	3	3	4	4	Public Transport	--	--	--	3

Note: In this table, p_50 represents percentile 50. The same goes to p_70, p_90, and p_100.
At the Flanders level, there is only one geographical zone, therefore the concept of p_50, p_70 and p_90 is not applicable.

4.2.4 Activity types

Concerning the activity-related index, the FEATHERS framework defines 10 different activity types. The results of four common activity types in our daily life are listed in Table 5. They are home-related activity, work-related activity, shopping activity, and touring activity, respectively. Regardless of the most stable geographical levels, i.e., the Flanders and Province levels, home-related activity needs less number of model runs to reach stability than that for work-related activity, which in turn requires less runs than that for shopping activity. Touring activity, however, requires the most model runs among these four types. Such an ordering appears to be quite consistent with the frequency of these activities happened in our daily life.

TABLE 5 The required minimum number of model runs for different percentiles of stable zones at 6 geographical levels by activity types on average daily number of activities per person.

BB (10521)	required minimum nr of runs				**Subzone (2386)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Home-related Activity	11	19	53	>200	Home-related Activity	5	7	14	>200
Work-related Activity	58	113	>200	>200	Work-related Activity	16	28	85	>200
Shopping Activity	175	>200	>200	>200	Shopping Activity	44	79	>200	>200
Touring Activity	>200	>200	>200	>200	Touring Activity	191	>200	>200	>200

Zone (1145)	required minimum nr of runs				**Superzone (327)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Home-related Activity	4	5	10	68	Home-related Activity	3	3	3	51
Work-related Activity	12	22	60	>200	Work-related Activity	5	6	7	>200
Shopping Activity	31	60	151	>200	Shopping Activity	9	11	15	>200
Touring Activity	133	>200	>200	>200	Touring Activity	28	40	57	>200
Province (6)	required minimum nr of runs				**Flanders (1)**	required minimum nr of runs			
	p_50	p_70	p_90	p_100		p_50	p_70	p_90	p_100
Home-related Activity	2	2	2	2	Home-related Activity	--	--	--	2
Work-related Activity	2	2	2	2	Work-related Activity	--	--	--	2
Shopping Activity	3	3	3	3	Shopping Activity	--	--	--	2
Touring Activity	3	3	3	3	Touring Activity	--	--	--	2

Note: In this table, p_50 represents percentile 50. The same goes to p_70, p_90, and p_100.
At the Flanders level, there is only one geographical zone, therefore the concept of p_50, p_70 and p_90 is not applicable.

4.3 Total vehicle kilometres travelled on the Flemish road network

As we can see from the above analysis, the whole Flanders, the highest level of geographic detail in this study, always reaches the predefined confidence interval fastest. Even taking the segmentations into account, limited number of FEATHERS runs are enough for the index under study to achieve the stable condition (i.e., $CI \le 0.1\times \overline{X}_{100}$). In this section, we investigate the variation of the vehicle kilometres travelled on the Flemish road network by carrying out traffic assignment 100 times. Specifically, after each model run, we obtain the predicted activity travel patterns of each agent, based on which the time-dependent origin-destination (OD) matrices (between 8 a.m. to 9 a.m.) can be derived. Afterwards, the traffic flow and the vehicle distance travelled on each network link can be simulated. Thus, at the whole Flanders level, the total vehicle kilometres travelled as well as the vehicle kilometres travelled for 8 different route link types can be computed. Now, based on the 100 traffic assignment results, the required minimum number of model runs can be identified. The results are shown in Table 6.

TABLE 6 The required minimum number of model runs at the whole Flanders level on vehicle kilometers travelled on the Flemish network.

Flanders	100 runs Average ($*10^5$ kilometer)	minimum nr of runs required (CI<=0.1*average value)
overall	72.312	2
Linktype1	30.063	2
Linktype2	1.596	3
Linktype3	18.991	3
Linktype4	0.071	7
Linktype5	9.036	3
Linktype6	1.985	3
Linktype7	9.343	3
Linktype8	1.228	2

Note: Linktypes1-8 represent the route link type of Flemish road network.

As can be seen, 2 FEATHERS runs are enough to meet the requirement of the confidence interval for the total vehicle kilometres travelled on the Flemish road network. Considering the 8 route link types (Linktypes1-8 in Table 6), highway (Linktype1) appears to be the most frequently used road type and is relatively easier to achieve the requirement. Nevertheless, even for the one with the lowest amount of vehicle kilometres travelled, i.e., Linktype4, 7 model runs would satisfy the condition. Such a result provides users with great confidence that application of the FEATHERS at an aggregated level only requires limited runs.

5 CONCLUSIONS

Activity-based models of travel demand provide outputs from micro-simulation-based forecasts. Therefore, stochastic errors due to the statistical distributions of random components are inherently included in such models. Analysis of their impacts on the model outputs thereby becomes one of the vital steps in the model development. In this study, the effect of stochastic error in the FEATHERS framework, an agent-based micro-simulation model particularly developed for Flanders, Belgium, was investigated, in which six levels of geographic detail were taken into account. The concept of confidence interval was applied with the purpose of determining the required minimum number of model runs to ensure at least a certain percentile of zones in each geographical level to reach their stability (i.e., with a certain level of confidence that the obtained average value of each of these zones can only vary within an acceptable interval).

By successively running the FEATHERS 100 times based on the 10% fraction of the full population, the variation of three travel indices including the average daily number of trips per person, the average daily distance travelled per person, and the average daily number of activities per person, as well as their corresponding segmentations with respect to socio-demographic variables (gender and age), transport mode choice, and activity types, were estimated. The results indicated a consistent phenomenon, i.e., for a percentile of zones, the index under study at a higher aggregated level was normally easier than at a lower level to achieve the predefined stable condition. Here, the degree of the aggregation referred to not only the size of the geographical scale, but also the detailed extent, i.e., the segmentation of the population, of the index under study.

Concerning the geographic scales, it needed only limited number of model runs at the highly aggregated levels (such as the Flanders and the province levels) to ensure all the zones (i.e., the 100 percentile) in these levels to be stable with respect to all the indices and their segmentations. By calculating the vehicle kilometers travelled on the Flemish road network after traffic assignment, similar results were obtained. All this provides model users with confidence that application of the FEATHERS at an aggregated level only requires limited model runs.

However, when it came to the BB level, the most disaggregated geographical level in this study, more than 200 model runs were usually required to enable all the zones to satisfy the stable condition for any index. And within 100 runs, normally only 70% or even 50% of the zones could be guaranteed in terms of their stability. It is therefore a dilemma between exploring in more complex or detailed circumstances and seeking for more reliable results. One compromising solution is to set another relatively achievable confidence interval condition for the zones with high variation, especially when these zones are not involved in the study area.

With regard to the different segmentations of population, we found that the required number of model runs was relatively fewer for the particular target segments which potentially involved more trips or activities. Specifically , the male group which generated a relatively larger number of trips and distance travelled needed a relatively fewer number of model runs than the female group to reach the predefined stability for each percentile of zones. Also, the required runs for the younger age categories (i.e., 18-34 years and 35-54 years) were apparently less than that for the other elder age categories (i.e., 55-64 years, 65-74 years, and over 75 years). Besides, the most frequently used transport mode in Belgium, i.e., the car as driver, compared to the others, required the fewest number of model runs to satisfy the predefined stable condition for both the trip and the distance related indices. Concerning the index of activity, home-related activity as the most frequently happened activities in our daily life needed less number of model runs to reach stability than the other activity types. Finally, the analysis of the vehicle kilometres travelled on the Flemish road network also showed similarly that the vehicle kilometres travelled on highways (Linktype1) were relatively easier to meet the requirement than that of other road types.

In the future, more aspects could be investigated. First of all, the impact of the population fraction on the stochastic error should be studied. New insights could probably be gained by repeating the experiment based on the full population instead of the 10% fraction. Moreover, based on the model outputs, other valuable travel indices could be taken into account as well, such as the index on travel time. In addition, apart from looking at the stochastic micro-simulation error in the FEATHERS, exploration on other potential uncertainty due to such as input variability and model specification is also worthwhile. Finally, it should be noticed that this study only focused on one modelling framework. Generalization of the findings to other activity-based travel demand models should therefore be a meaningful future research direction.

REFERENCES

(1) McNally, M.G. *The four step model*. In: Hensher, D.A. and K.J. Button (Eds.), Handbook of Transport Modelling, Elsevier Science, Oxford, UK, 2007.

(2) Jones, P.M., F.S. Koppelman, and J.P. Orfueil. *Activity analysis: State-of-the-art and future directions*. In: Jones, P.M. (Ed.), New Developments in Dynamic and Activity-based Approaches to Travel Analysis, Gower Publishing, Aldershot, England, 1990, pp. 34-55.

(3) Rasouli, S., and H.J.P. Timmermans. Uncertainty in travel demand forecasting models: literature review and research agenda. *Transportation Letters: The International Journal of Transportation Research*, Vol. 4, 2012, pp. 55-73.

(4) Smith, L., R. Bechman, K. Baggerly, D. Anson, and M. Williams. TRANSIMS: Transportation Analysis and Simulation System: Project Summary and Status, 1995.

(5) Veldhuisen, K., H.J.P. Timmermans, and L.L. Kapoen. Ramblas: A regional planning model based on the micro-simulation of daily activity travel patterns, *Environment and Planning A*, Vol. 32, 2000, pp. 427-443.

(6) Bhat, C.R., J.Y. Guo, S. Srinivasan, and A. Sivakumar. A comprehensive microsimulator for daily activity-travel patterns. In *Proceedings of the Conference on Progress in Activity-Based Models*, Maastricht, CD-ROM, 2004.

(7) Pendyala, R.M., R. Kitamura, A. Kikuchi, T. Yamamoto, and S. Fujji. FAMOS: Florida activity Mobility simulator. In *Proceedings of the 84th Annual Meeting of the Transportation Research Board*, Washington D.C., CD-ROM, 2005.

(8) Arentze, T.A., and H.J.P. Timmermans. *Albatross: A Learning-Based Transportation Oriented Simulation System.* European Institute of Retailing and Services Studies. Eindhoven, the Netherlands, 2000.

(9) Arentze, T.A., F. Hofman, H. van Mourik, and H. Timmermans. ALBATROSS: Multiagent, Rule-Based Model of Activity Pattern Decisions. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1706, Transportation Research Board of the National Academies, Washington, D.C., 2000, pp. 136-144.

(10) Arentze, T.A., and H.J.P. Timmermans. A Learning-based Transportation Oriented Simulation System. *Transportation Research Part B: Methodological*, Vol. 38, No. 7, 2004, pp. 613-633.

(11) Bellemans, T., D. Janssens, G. Wets, T.A. Arentze, and H.J.P. Timmermans. Implementation Framework and Development Trajectory of Feathers Activity-Based Simulation Platform. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2175, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 111-119.

(12) Arentze, T., H.J.P. Timmermans, D. Janssens, and G. Wets. Modeling Short-Term Dynamics in Activity-Travel Patterns: From Aurora to Feathers. In *Transportation Research Record Conference Proceedings 42*, Vol. 2, 2008, pp. 71-77.

(13) Timmermans, H.J.P., T.A. Arentze, and C-H. Joh. Analysing Space-time Behavior: New Approaches to Old Problems. *Progress in Human Geography*, Vol. 26, No. 2, 2002, pp. 175-190.

(14) Castiglione, J., J. Freedman, and M. Bradley. Systematic Investigation of Variability due to Random Simulation Error in An Activity-based Micro-simulation Forecasting Model. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1831, Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 76-88.

(15) Kochan, B., T. Bellemans, D. Janssens, and G. Wets. Assessing the Impact of Fuel Cost on Traffic Demand in Flanders using Activity-based Models. In *Proceedings of Travel Demand Management*, Vienna, Austria, 2008.

(16) Knapen, L., B. Kochan, T. Bellemans, D. Janssens, and G. Wets. Activity-based modeling to predict spatial and temporal power demand of electric vehicles in Flanders, Belgium, In *Transportation Research Record: Journal of the Transportation Research Board*, No. 2287, Transportation Research Board of the National Academies, Washington, D.C., 2012, pp. 146-154.

(17) Kusumastuti, D., E. Hannes, D. Janssens, G. Wets, and B.G.C. Dellaert. Scrutinizing Individuals' Leisure-shopping Travel Decisions to Appraise Activity-based Models of Travel Demand. *Transportation*, Vol. 37, 2010, pp. 647-661.

(18) Dowling, R., A. Skabardonis, and V. Alexiadis. *Traffic Analysis Toolbox Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software*. Publication FHWA-HRT-04-040. FHWA, U.S. Department of Transportation, 2004.

(19) Cools, M., B. Kochan, T. Bellemans, D. Janssens, and G. Wets. Assessment of the Effect of Microsimulation Error on Key Travel Indices: Evidence from the Activity-based Model Feathers. In *Proceedings of the 90th Annual Meeting of the Transportation Research Board*, Washington D.C., CD-ROM, 2011.