Research on Restrained Study Areas for Effective Activity-based Travel Demand Forecasting

Qiong BAO[1], Yongjun SHEN[1], Bruno KOCHAN[1], Tom BELLEMANS[1], Davy JANSSENS[1], and Geert WETS[1]

[1] Transportation Research Institute (IMOB), Hasselt University, Wetenschapspark 5 bus 6, 3590, Diepenbeek, Belgium; PH +32 11 269146; FAX +32 11 269199; E-Mail: {qiong.bao, yongjun.shen, bruno.kochan, tom.bellemans, davy.janssens, geert.wets}@uhasselt.be

ABSTRACT

Recently, considerable attention has been devoted to studying the activity-based approach for transportation planning and forecasting. However, one of the practical limitations of applying most of the currently available activity-based models is their computation time. In this research, we investigated the possibility of restraining the size of the study area to reduce the computation time when applying an activity-based model. By introducing an accuracy level of the model, we proposed an iterative approach to determine the minimum size of the study area required for a target territory. In the application, we investigated the required minimum size of the study area surrounding each of the 327 municipalities in Flanders with regard to two different transport modes: car as driver and public transport. Additionally, a validation analysis was conducted. All the experiments were carried out by using the FEATHERS (Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS) framework, an activity-based micro-simulation modeling framework currently implemented for the Flanders region of Belgium.

INTRODUCTION

As an alternative to the traditional four-step model of travel demand, the activity-based approach has recently been the focus of many transportation researchers and has resulted in the development of a number of practical models such as ALBATROSS (Arentze et al., 2000), RAMBLAS (Veldhuisen et al., 2000), CEMDAP (Bhat et al., 2004), and FAMOS (Pendyala et al., 2004). Since the approach focuses on the complete activity behavior patterns and adopts a holistic framework considering the individual interactions and spatio-temporal constraints, it explicitly reveals the inability of the conventional trip-based approach. It also can be

used to address many policy issues and their impact such as land use, energy consumption, emission, safety, and congestion pricing (Jones et al., 1990; McNally, 2007; Malayath and Verma, 2013).

Although its usefulness in transportation planning and forecasting has been widely recognized, one of the practical limitations of applying most of the currently available activity-based models is their computation time, especially when a large population and detailed geographical unit level are considered. For instance, in the FEATHERS (Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS) framework (Bellemans et al., 2010), an activity-based micro-simulation modeling framework currently implemented for the Flanders region of Belgium, it takes approximately 16 hours for a single model run in which a simple case is analyzed. In this scenario, the model is based on the 10% of the full population of Flanders at the building block level (currently the most disaggregated geographical level of detail for Belgium) (Bao et al., 2013). If the population fraction increases to 50%, a completely feasible scenario commonly used in the model operation, the FEATHERS framework will then take almost two days to complete the model execution. Moreover, if multiple model runs are required due to the consideration of stochastic error derived from the micro-simulation approach (Bao et al., 2013), the computation time will be magnified dramatically, which makes the real-time application of the model particularly difficult or even impossible to realize.

In order to reduce the computation time when applying an activity-based model, several tradeoffs can be made in actual applications. One such possibility is to restrain the size of the study area and to conduct the computation only for the selected region. In many cases, this may be practical as it is often the case that merely a small territory (e.g., a municipality) rather than the whole region or country is the focus of a specific study. Therefore, a relatively small study area surrounding the target territory may only be needed for investigation rather than taking the whole region into account. In cases like this, the computation time of the model could be drastically reduced. The question then becomes: what should be the minimum size of the study area surrounding the target territory and how does one determine it? Based on authors' current knowledge and literature search, however, few studies have investigated this important issue to date. In most of the current studies using activity-based models, the size of the study area is chosen mainly via domain knowledge of researchers or practitioners themselves. This choice is often rather arbitrary. In this study, by defining an accuracy level of the model, we propose an iterative approach to determine the minimum size of the study area required for a target territory when performing travel demand forecasting. More specifically, by adding a small zone to the target territory constituting a new study area each time,

the accuracy of the model is calculated. In this paper, accuracy is defined as the difference between the occurrence of both the departing and the arriving trips derived based on this study area and that based on the whole. Such a procedure is repeated until the predefined accuracy level is satisfied. In the application, we investigate the required minimum size of the study area surrounding each of the 327 municipalities in Flanders, Belgium with regard to two different transport modes: car as driver and public transport. Afterwards, a validation analysis based on four extreme municipalities is conducted. All the experiments are carried out by using the FEATHERS framework.
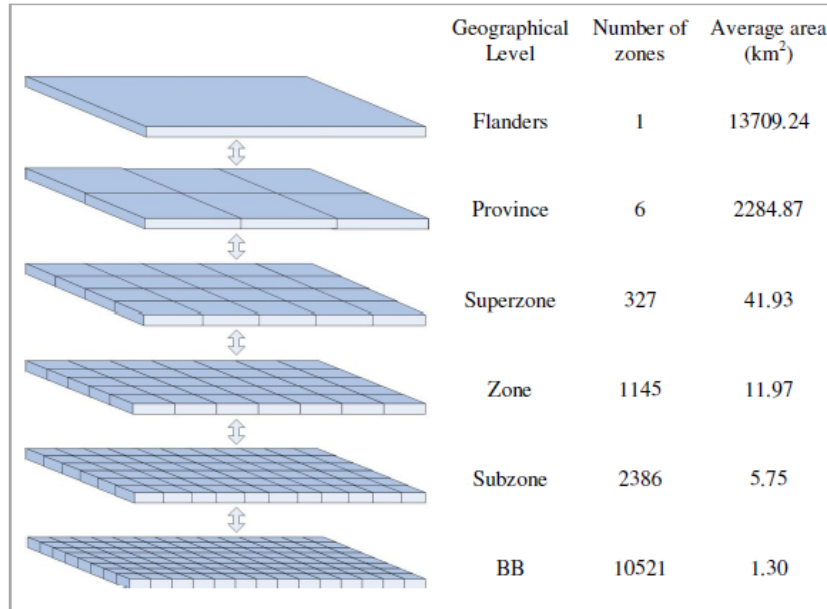
In the remainder of the paper, we are by briefly introducing the FEATHERS framework and the levels of geographic detail of Flanders. Next, the methodology proposed in this research to determine the minimum size of the study area is established, followed by a detailed demonstration of the experiment execution. Subsequently, the results are presented and discussed. Finally, the paper ends with conclusions and some ideas for future research.

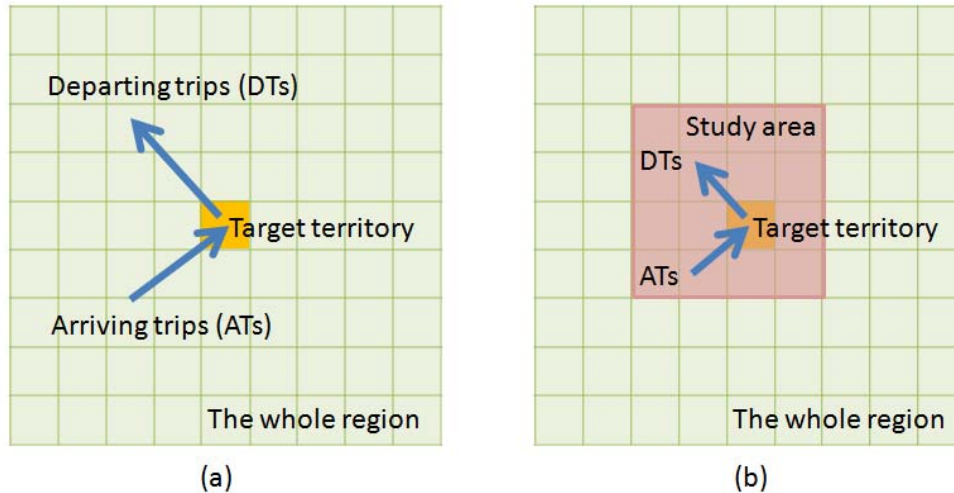## FEATHERS FRAMEWORK FOR FLANDERS

The FEATHERS (Forecasting Evolutionary Activity-Travel of Households and their Environmental RepercussionS) framework (Bellemans et al., 2010) is a micro-simulation framework developed to facilitate the implementation of activity-based models for transport demand forecasting. Currently, an activity-based model similar to the ALBATROSS model (Arentze et al., 2008) is embedded. In this model, 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. These attributes include characteristics 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). For each individual person and his/her 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 has been implemented for the Flanders region of Belgium (e.g., Kochan et al., 2008; Kusumastuti et al., 2010; Knapen et al., 2012) and is fully operational at six levels of geographic detail of Flanders. These six levels include the Building block (BB) level, Subzone level, Zone level, Superzone level, Province level, and the whole Flanders level. Figure 1 illustrates the hierarchy of the geographical layers with different granularities.

In practice, to predict the travel demand or the total number of trips to occur within a specific zone of Flanders (i.e., a target territory, we normally have to calculate both the departing trips (i.e., the trips from this target territory to the whole Flanders region) and the arriving trips (i.e., the trips from the whole Flanders region to this target territory). Figure 2 shows schematics that describe the calculations. consequence of the model is that the more detailed the geographical unit level considered, the longer the computation time needed. For instance, to run FEATHERS at the Subzone level, approximately 16 hours are needed based on the 50% of the full population of Flanders. If the most disaggregated geographical level of detail, i.e., the BB level is under consideration, the FEATHERS framework will then take almost two days to complete the model execution. Therefore, determining how to effectively reduce the model computation time is a practical issue of applying this 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.

Figure 2. The schematic diagrams of travel demand calculation for a target territory.

METHODOLOGY AND EXPERIMENT

In this study, we aim to find an effective solution to the computation time problem of activity-based models in general, and for the FEATHERS framework in particular. As described in the previous section, to estimate the total number of trips occurring within a target territory, the entire region of Flanders is normally used as the study area to calculate both the departing and arriving trips of this territory. However, if we can find a relatively small study area surrounding this target territory within which most of the departing and arriving trips are generated, such as that shown in Figure 2(b), it is then not necessary to take the whole Flanders region into account. In this case, the computation time of the model could be dramatically reduced. The question then becomes: what should be the minimum size of the study area surrounding the target territory and how does one determine it?

In this research, we propose an iterative approach based on which we investigate the minimum size of the study area needed for each of the 327 municipalities (i.e., the Superzone level) in Flanders, Belgium. The whole procedure is illustrated in Figure 3. More specifically, by generating the basic prediction dataset from the activity-based model inside FEATHERS, we obtain the whole activity travel pattern or schedule information for each individual territory in Flanders, based on which the origin and destination (OD) matrices can be derived.
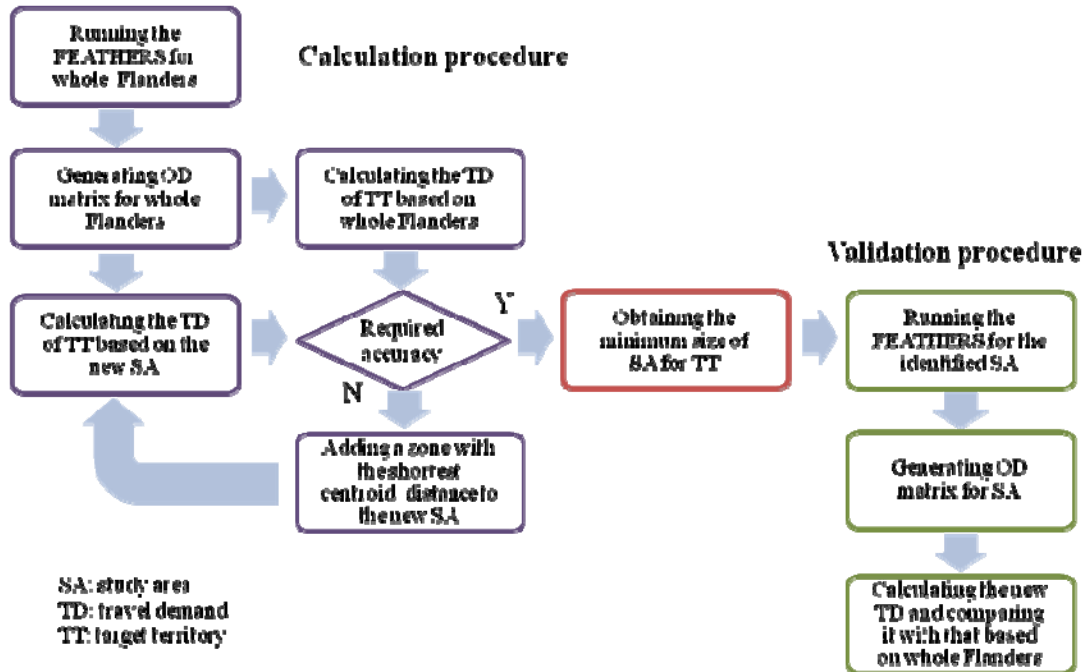
Figure 3. The flow diagram of the calculation and the validation procedure.

Next, for each particular Superzone *i*, one more zone *j* (which could be a Superzone, a Subzone, or a Building block) with the shortest centroid distance to the target Superzone *i* is added constituting a new study area (SA). Then, the travel demand (i.e., the number of trips) of both the departing mode and the arriving mode within this new study area is computed. These quantities can then be used to calculate the difference (D_{ij}) for the whole Flanders region via Equation 1, Equation 2 can then be used to estimate the accuracy rate.

$$D_{ij} = \frac{\sqrt{(\#DTs\ in\ SA_{ij} - \#DTs\ in\ Flanders)^2 + (\#ATs\ in\ SA_{ij} - \#ATs\ in\ Flanders)^2}}{\sqrt{(\#DTs\ in\ Flanders)^2 + (\#ATs\ in\ Flanders)^2}} \quad (1)$$

$$A_{ij} = (1 - D_{ij}) \times 100\% \quad (2)$$

Such a procedure is repeated by adding more zones until the predefined accuracy requirement is reached. When zones are added, those nearest the target territory are added first, followed by those at increasing distances from the target territory. The study area obtained through this method is thus the minimum size needed for the travel demand prediction of the target municipality. Under this process, the centroid distance between the last zone that was added into the study area and the target municipality is defined as the minimum radius of the study area surrounding this municipality. Here, the radius is interpreted differently than in its conventional definition, but refers in particular to the centroid distance between the

added zone and the target municipality. Therefore, the radius for each municipality increases discretely, and when the radius increases by one unit, only one zone is counted into the study area.
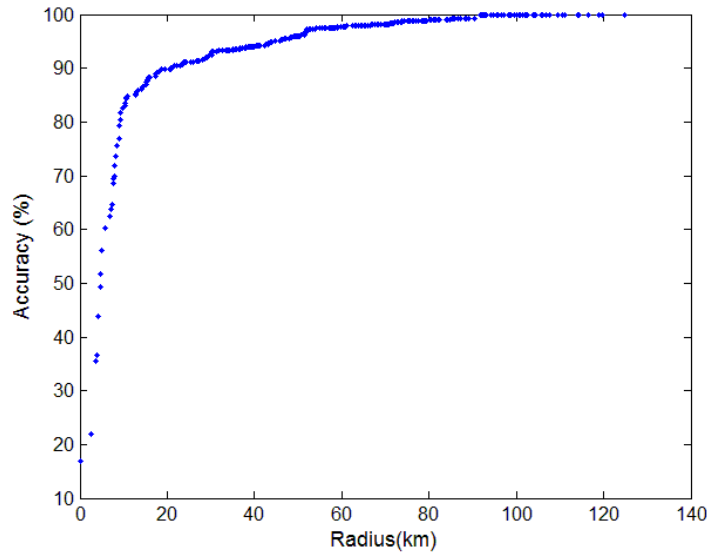
In the experiments, the FEATHERS framework is executed at the Subzone level for the 50% fraction of the full population, and 90% accuracy level is selected. Under these conditions, we investigate the minimum size of the study area needed for each of the 327 municipalities in Flanders, Belgium, with regard to two different transport modes: car as driver and public transport, respectively. Moreover, as a validation procedure (detailed in Figure 3), we run the FEATHERS framework again based on the identified study area respectively for four extreme cases including the municipality with the longest study area radius and the one with the shortest for each mode. The accuracy of the model in each case is examined, and the degree of the reduction in computation time is estimated.

RESULTS AND DISCUSSION

By applying the aforementioned methodology, the corresponding results are presented and further discussed in the following sections.
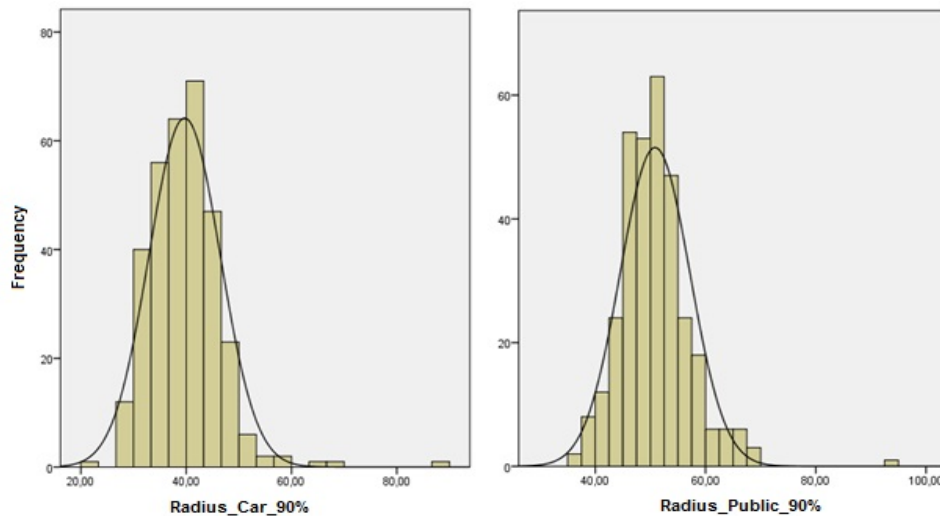
The minimum study area required for 327 municipalities in Flanders

In the experiment the radius of the study area is increased each time surrounding municipalities are added one by one, in a process starting with those nearest the target territory and moving away. As a result, the difference of the travel demand regarding the departing mode and the arriving mode between the current study area and the whole Flanders region, as calculated via Equation 1,is expected to decrease gradually. The accuracy level of the new study area, calculated via Equation 2, will increase correspondingly. In taking a randomly selected municipality as an example, the relationship between the achieved accuracy and the required radius of the study area is presented in Figure 4.

Figure 4. The relation between the accuracy level and the required radius for a randomly selected municipality.

To further obtain the minimum radius of the study area needed for each of the 327 municipalities in Flanders, a 90% accuracy level is selected in this study with respect to two different transport modes: car as driver and public transport. The distribution of the results for all the 327 municipalities is illustrated in Figure 5.
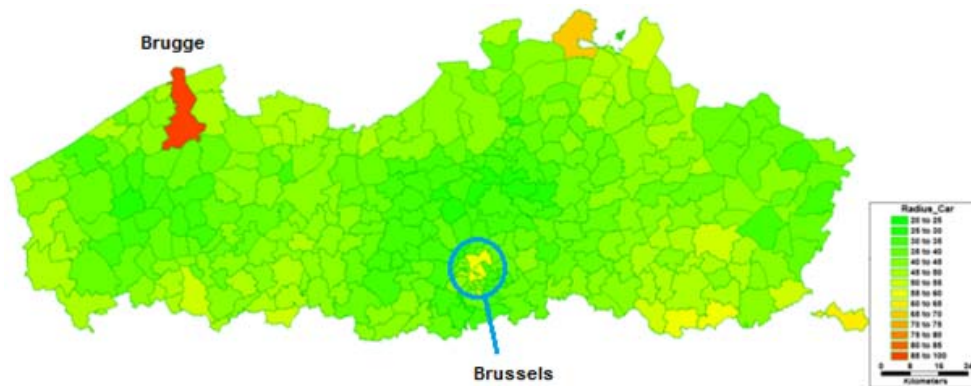

Figure 5. The distribution of the minimum radius needed for each of the 327 municipalities in Flanders with respect to car as driver mode and public transport mode.

The figure shows that the required radius for both the modes follows the normal distribution, and the average of the minimum radius needed for the car as driver mode is 39.70 km with a standard deviation of 6.77 km. The average radius for the public transport mode is 50.84 km with a standard deviation of 6.33 km. In

other words, to achieve the same 90% accuracy level for each municipality, the public transport mode generally needs a relatively larger study area compared to the car as driver mode. This can be partly explained by the fact that people in Flanders are more likely to choose the public transport mode (e.g., train) for a long distance trip, especially when the distance is larger than 50 km. Even so, given the fact that the area of the whole Flanders region is 13709.24 km^2, the size of the study area needed for most of the municipalities is reduced to a great extent for both the transport modes, and therefore the computation time of the FEATHERS is drastically reduced.
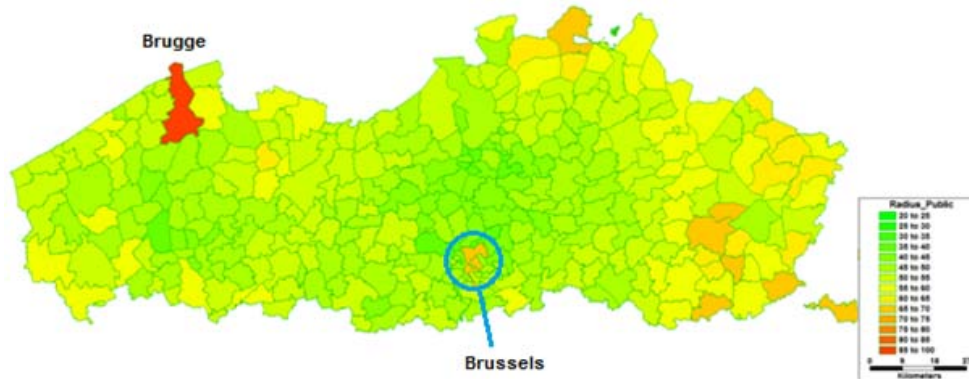
To more clearly show the results, we present the calculated minimum radius of the study area for each of the 327 municipalities visually by using a color theme with 14 different colors for the car as driver mode and the public transport mode, respectively. The results are shown in Figures 6 and 7, in which the displayed color tends toward green when a shorter radius is needed for a municipality, while the displayed color tends toward red when the radius needed is becoming larger.

Figure 6. The color theme of the minimum radius needed for 327 municipalities of Flanders with respect to car as driver mode at 90% accuracy level.

Based on these two figures, we can clearly see that for most of the municipalities of Flanders, the required minimum radius (or study area) is typically larger for the public transport mode than that for the car as driver mode. In other words, the computation time of the FEATHERS framework will be reduced more drastically when the car as driver mode is under consideration. Moreover, for both the modes, especially the public transport mode, the municipalities lying on the border of Flanders generally need larger radii to reach the given accuracy level compared to those located in the relatively central position of Flanders. For instance, the City of Brugge – one of the leading seaside resorts in Belgium – requires the largest radius, regardless of which transport mode is considered. In addition, the capital city Brussels, which served as a traffic hub of the region, also needs a larger study area for both the modes compared to its neighboring municipalities. This is
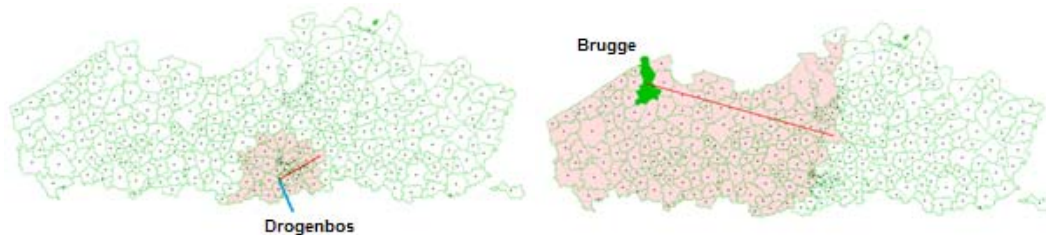
due to the fact that a great number of trips are happening every day between this municipality and all the others.

Figure 7. The color theme of the minimum radius needed for 327 municipalities of Flanders with respect to public transport mode at 90% accuracy level.

Extreme cases in Flanders based on the identified minimum study area

To verify the results we obtained, the validation procedure shown in Figure 3 is conducted by considering the four extreme cases: the municipality with the longest study area radius and the one with the shortest for both the car as driver mode and the public transport mode (see Figures 8 and 9).

Figure 8. The municipality with the shortest and the longest required radius of study area with respect to car as driver mode at 90% accuracy level.

Figure 9. The municipality with the shortest and the longest required radius of study area with respect to public transport mode at 90% accuracy level.

By running the FEATHERS framework again based on the identified study area for each extreme case, we can predict the travel pattern of each individual, from

which the origin and destination matrix in this territory can be derived. Then, by calculating the new travel demand of both the departing mode and arriving mode within this study area and further comparing it with the one based on the whole Flanders region, we can examine the achieved accuracy level of the model. More importantly, by recording the computation time for each extreme case, the degree of the time reduction can be estimated. The results are shown in Table 1.

Table 1. The validation results for the extreme cases with respect to car as driver mode and public transport mode

Traffic mode	Municipality	Radius		Accuracy level	Running Time (hour)	Time saving
Car as driver	Drogenbos	Shortest:	21.2 km	86.9%	4	75%
	Brugge	Longest:	89.82 km	87.1%	9	43.8%
Public transport	Wemmel	Shortest:	36.17 km	86.3%	7.5	53.1%
	Brugge	Longest:	94.24 km	85.4%	10.5	34.4%

After performing the travel demand forecasting based on the rebuilt study area, we can see that all four extreme cases show a very high accuracy rate (all above 85%). Moreover, by restraining the size of the study area, the computation time of the model is reduced dramatically, from 16 hours in the original case to as little as 4 hours (i.e., a maximum of 75% time savings). We can therefore conclude that running the activity-based model inside the FEATHERS framework using the rebuilt study area for each municipality of Flanders will improve the model's operational efficiency significantly.

CONCLUSIONS AND FUTURE RESEARCH

The requirement of large computation time is currently one of the most important practical issues of applying activity-based models for travel demand forecasting. In this study, we investigated the possibility of restraining the size of the study area in order to reduce the computation time, as in many cases, only a small territory rather than the whole region is the focus of a specific study. By introducing an accuracy level of the model, which is defined as the difference between the occurrence of both the departing and the arriving trips derived based on the study area and that based on the whole region, we proposed an iterative approach to determine the minimum size of the study area needed for a target territory. Additionally, both the calculation and validation procedures for this approach were designed. In the application, all of the 327 municipalities of Flanders, Belgium were

studied using the FEATHERS framework, an activity-based micro-simulation modeling framework. The minimum size of the study area needed for each of these municipalities was computed with regard to the car as driver mode and the public transport mode, respectively, under a predefined accuracy level of 90%. The results indicated that the municipalities lying on the border of Flanders (e.g., the City of Bruges) as well as some traffic hub cities (e.g., the capital city Brussels) generally need larger radii to reach the given accuracy level. Meanwhile, for most of the municipalities of Flanders, the required minimum radius (or study area) is typically larger for the public transport mode than that for the car as driver mode.

To verify the results we obtained, a validation analysis was carried out by running the FEATHERS framework based on the identified study area for four extreme cases: the municipality with the longest study area radius and the one with the shortest for both the car as driver mode and the public transport mode. It turned out that within the identified minimum size of each study area, the computation time of the FEATHERS framework was reduced considerably (up to 75%), while the model still reached a very high accuracy rate. Such a result confirmed that when only a particular territory is needed for consideration in a specific study, it is possible to rebuild a relatively small study area for investigation. Running the model in such a restrained study area will improve the model's operational efficiency.

In the future, the results obtained in this paper can be consulted as a reference for those who plan to use the FEATHERS framework. For other activity-based models, the methodology proposed in this paper for the calculation of a minimum size of study area can also be repeated. However, more aspects need to be investigated as well. First, other accuracy levels can be considered, and the best tradeoff between the accuracy rate and the computation time can be discussed. Moreover, apart from analyzing the results based on different transport modes, other valuable travel indices could be taken into account as well, such as activity types. In addition, exploration on detailed reasons behind the different size of the study area needed for each target territory is also worthwhile, which will in turn validate this modeling framework and facilitate its further development.

REFERENCES

Arentze, T.A., Hofman, F., van Mourik, H., and Timmermans, H.J.P. (2000). "ALBATROSS: Multiagent, Rule-Based Model of Activity Pattern Decisions." *Transportation Research Record*, *1706*, Transportation Research Board, Washington, D.C., 136-144.

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

Bao, Q., Kochan, B., Bellemans, T., Janssens, D., and Wets, G. (2013). "Investigating micro-simulation error in activity-based travel demand forecasting using confidence intervals." *Proceedings of the 92nd Annual Meeting of the Transportation Research Board* (Compendium of Papers, CD-ROM), No. 13-2528.

Bellemans, T., Janssens, D., Wets, G., Arentze, T.A., and Timmermans, H.J.P. (2010). "Implementation Framework and Development Trajectory of Feathers Activity-Based Simulation Platform." *Transportation Research Record*, *2175*, Transportation Research Board, Washington, D.C., 111-119.

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

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

Knapen, L., Kochan, B., Bellemans, T., Janssens, D., and Wets, G. (2012). "Activity-based modeling to predict spatial and temporal power demand of electric vehicles in Flanders, Belgium." *Transportation Research Record*, *2287*, Transportation Research Board, Washington, D.C., 146-154.

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

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

Malayath, M. and Verma, A. (2013). "Activity based travel demand models as a tool for evaluating sustainable transportation policies." *Research in Transportation Economics*, 38, 45-66.

McNally, M.G. (2007). "The activity-based approach." D.A. Hensher and K.J. Button, eds., *Handbook of Transport Modelling*, Elsevier Science, Oxford, UK, 53-70.

Pendyala, R.M., Kitamura, R., and Kikuchi, A. (2004). "FAMOS: the Florida activity mobility simulator." *Proceedings of the Conference on Progress in Activity-Based Models*, Maastricht, CD-ROM.

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