# Accepted Manuscript

Activity-based model for medium-sized cities considering external activity-travel: Enhancing FEATHERS framework

Syed Fazal Abbas Baqueri, Muhammad Adnan, Bruno Kochan, Tom Bellemans

PII: S0167-739X(18)31431-6
DOI: https://doi.org/10.1016/j.future.2019.01.055
Reference: FUTURE 4748

To appear in: *Future Generation Computer Systems*

Received date : 3 October 2018
Revised date : 24 January 2019
Accepted date : 26 January 2019

# Activity-based Model for Medium-Sized Cities Considering External Activity-Travel: Enhancing FEATHERS Framework

Syed Fazal Abbas Baqueri[a,b], Muhammad Adnan[1,a], Bruno Kochan[a] & Tom Bellemans[a]

[a]UHasselt - Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium
[b]Civil Engineering Department, DHA Suffa University, Karachi 75500, Pakistan

## Abstract

Travel demand modeling has evolved from the traditional four-step models to tour-based models which eventually became the basis of the advanced Activity-Based Models (ABM). The added value of the ABM over others is its ability to test various policy scenarios by considering the complete activity-travel pattern of individuals living in the region. However, the majority of the ABM restricts residents' activities within the study area which results in distorted travel patterns. The external travel is modeled separately via external models which are insensitive to policy tests that an ABM is capable of analyzing. Consequently, to minimize external travel, transport planners tend to define a large study area. This approach, however, requires huge resources which significantly deterred the worldwide penetration of ABM. To overcome these limitations, this study presents a framework to model residents' travel and activities outside the study area as part of the complete activity-travel schedule. This is realized by including the Catchment Area (CA), a region outside the study area, in the destination choice models. Within the destination choice models, a top-level model is introduced that specifies for each activity its destination inside or outside the study area. For activities to be performed inside the study area, the detailed land use information is utilized to determine the exact location. However, for activities in the CA, another series of models are presented that use land use information obtained from open-source platforms in order to minimize the data collection efforts. These modifications are implemented in FEATHERS, an ABM operational for Flanders, Belgium and the methodology is tested on three medium-sized regions within Flanders. The results indicate improvements in the model outputs by defining medium-sized regions as study areas as compared to defining a large study area. Furthermore, the Points of Interests (POI) density is also found to be significant in many cases. Lastly, a comprehensive validation framework is presented to compare the results of the ABM for the medium-sized regions against the ABM for Flanders. The validation includes the (dis)aggregate distribution of activities, trips, and tours in time, space and structure (e.g. transport modes used and types of activities performed) through eleven measures. The results demonstrate similar distributions between the two ABM (i.e. ABM for medium-sized regions and for Flanders) and thus confirms the validity of the proposed methodology. This study, therefore, shall lead to the development of ABM for medium-sized regions.

**Keywords:** Activity-based Model, External Activity-Travel, External trips, FEATHERS, Activity-based model validation.

## 1. Introduction

The notion that the need for activity participation derives its associated travel, led to the formation of the Activity Based Model (ABM) (Ben-akiva et al., 1996). A typical ABM considers the complete daily

[1] Corresponding Author

E-mail addresses: fazal.abbas@dsu.edu.pk (Syed Fazal Abbas Baqueri), muhammad.adnan@uhasselt.be (Muhammad Adnan), bruno.kochan@uhasselt.be (Bruno Kochan), tom.bellemans@uhasselt.be (Tom Bellemans)

activity-travel pattern of individuals living in the study area. This includes, for each agent in the synthetic population, the number of activities to be performed and specific attributes of each activity: type, start time, duration, and location. Furthermore, these simulated activities are also linked together via a travel component having its own dimensions: travel time, travel mode. Finally, the tours are formed. Therefore, the added benefits of an ABM over a four-step model are its unit of analysis from zones to individuals and the consistency between the submodels that ensures a consistent travel pattern.
In reality, subjected to the attractiveness of the study area and its surrounding region, some of the activities can be performed outside the study area which results in residents Internal-External (IE) trips. However, the majority of the ABM does not model IE trips as they allow the destination choice of activities only within the modeling region, e.g. DAYSIM, ABM within SimMobility (Singapore) and FEATHERS (Flanders, Belgium, and Seoul, South Korea) (Adnan, Pereira, Miguel, et al., 2016; Bellemans et al., 2010; Bowman & Bradley, 2006). The ABM output is fed in the route assignment along with internal-external trips obtained from other models. Such an approach may result in the following deficiencies:

- Overestimating trips and activities within the region by assigning all residents' activities within the study area while completely disregarding the residents' external activities and trips.
- A *double representation* of residents' external trips at the route assignment stage, i.e. 1) from the ABM where external activity-travel of individuals is considered as internal trips and 2) through the output from the external trips model.
- Inability to test policy applications on resident's external travel because these are estimated outside the scope of the ABM.

These limitations have been well recognized and to overcome them, modelers tend to define a more extensive study area. Although this practice may reduce overall external travel, it increases the data collection and model development efforts: collecting household travel survey (HTS) data for a larger study area, preparing its synthetic population and running the ABM.

Consider a case of East Hampshire District Council (EHDC) - a medium-size district in the South East Region of England approximately 100km away from London. Expectedly, a lot of individuals commute from EHDC to London. Therefore, a travel demand model for EHDC should also include East of England and London Regions (formally government office region) in the study area (as recommended in Department for Transport 2017, p.13). This expansion of the study area results in unwanted model complexities such as modeling the travel behavior of Londoners which is indeed not the central objective. Likewise, expanding the study area may not always be a solution because of for example a boundary between two countries, resulting in data collection issues. For instance, the present ABM for Singapore (Siyu, 2015) is subjected to this issue as it assigns the residents' activities within Singapore, whereas, a lot of individuals frequently travel to Malaysia. As a result, the resident/s trips are over assigned within Singapore while completely ignoring their external travel. Detailed practical examples of these limitations are defined in Baqueri et al. (2018).

Consequently, only a few ABMs are operational at present mainly subjected to huge data collection and resources. Whereas, in order to develop a travel demand model for a medium-sized region, modelers have to rely on conventional four-step models. Therefore, it can be safely stated that the ability to model residents' external travel within ABM shall pave the way to develop an ABM for a medium-sized region. In light of these concerns, this study presents a framework to model residents external trips in FEATHERS - an activity-based travel demand model (Bellemans et al., 2010). The framework includes 1) defining an external region as Catchment Area (CA) within the ABM and 2) inclusion of CA within destination choice set. To limit the data collection efforts, the land use information of the CA is solely obtained using the open-source information to minimize the data collection cost. The study also describes the application of the proposed framework in three medium-sized study areas in Flanders, Belgium. Furthermore, a validation framework for ABM along with its implementation is also presented to compare the results of the proposed model against the model without a CA.

The rest of the paper is arranged as follows. The next section summarizes the literature on modeling external travel within ABM and ABM validation. The third section describes the modified FEATHERS framework. The fourth section describes the case study: the implementation study areas and the model results for each. The fifth section describes a framework for model validation along with aggregate and disaggregate validation. The sixth section provides a discussion of results and validation and the last section presents the conclusion.

# 2. Literature review

### 2.1 Activity-Based Model

Since their inception, the activity-based models have achieved significant progress in terms of theory, implementation, and deployment. Researchers and practitioners, particularly in the USA, Europe, and Japan develop and implemented ABMs. CARLA (constraint-based), STARCHILD (Recker et al., 1986a; Recker et al., 1986b), SCHEDULER (Gärling et al., 1994), DAYSIM (Bowman & Ben-Akiva, 1998), TRANSIMS (Smith et al., 1995), and ALBATROSS (Arentze & Timmermans, 2004) are some early examples of the ABM (Siyu, 2015, p.14).

ADAPTS (Agent-based Dynamic Activity Planning and Travel Scheduling), TASHA (Travel/Activity Scheduler for Household Agents) and SimMobility are some advanced prototypes of the ABM. These ABMs have much more sophisticated model structure to deal with the complex transport system (Auld & Mohammadian, 2012; Miller & Roorda, 2003; Adnan, Pereira, Miguel, et al., 2016). For instance, unlike other ABM frameworks, ADAPTS have an activity planning step that incrementally plans and updates activities for each individual for each time interval. TASHA models, for each individual in a household, its vehicle allocation, ridesharing and joint activities/trips. SimMobility integrates long-term models such as vehicle ownership, land use pattern with daily schedule and within day rescheduling such as disruption strategies. It also includes mode and destination accessibility for each individual through logsums.

With the passage of time, the spectrum of ABM has been constantly expanding to more advanced issues such as the demand for electric vehicles charging stations (Usman et al., 2017), Disruption Management Strategies (Adnan, Pereira, Azevedo, et al., 2016), carpooling demand (Hussain et al., 2016) and integration of autonomous vehicles in ABM (Childress et al., 2015). Recently, ABM has also demonstrated its multidisciplinary potential such as linking transportation with air quality analysis (Shabanpour et al., 2016), traffic noise (Kaddoura et al., 2017), energy demand and power-peaks (Weiss et al., 2017; Knapen et al., 2012), emissions and environmental impacts (Shiftan et al., 2015), and health assessments (Lefebvre et al., 2013). Therefore, it can be well guessed that the ABM will continue to maintain their impetus in future as well.

At present, most of the ABM disregard external travel and estimate them unconnectedly through other external models. The external trip models are analogues to first two steps of the four-step model as they predict aggregate external trip generation at external stations, i.e., highway intersections at the boundary of the study area and distribute them in the Traffic Analysis Zones (TAZs) of the study area. The travel mode for external trips is not explicitly modeled as usually cars are considered as travel mode and the OD matrix is directly used for route assignment along with the results of the ABM. Such an approach results in numerous problems as described in the previous section. However, few ABMs do consider the outside area through the additional zone(s) in the destination choice model. For example, ALBATROSS considers the surrounding area as one additional zone (Arentze & Timmermans, 2004). Similarly, ADAPTS – a state-of-the-art ABM, assigns external destinations to several zones around the Chicago region (Auld & Mohammadian, 2012). However, the size of these external zones is very large as compared to the zones within the study area. Due to this, travel times and cost of trips between the study area and the surrounding region will be inappropriate and, therefore, sub-models within ABM that requires these inputs may not perform well. To address these stated concerns, this paper presents a comprehensive framework that includes the residents' external travel within the ABM framework.

### 2.2 Activity-Based Model Validation

Model validation is an important aspect. However, there are limited studies that describe validation of travel demand models (de Jong et al., 2007; Rasouli & Timmermans, 2012). The studies vary according to the type of the model (rule based, utility based), aggregation level and uncertainty analyzed. Many studies described ABM validation by focusing on the discrete choice models (Castiglione et al., 2003; Gibb & Bowman, 2007; Bekhor et al., 2014) or a rule based approach (Zhuge et al., 2017; Cools et al., 2011; Bao et al., 2015; Bao et al., 2016; Rasouli, 2016). Majority of the studies focus on the core activity-scheduling part (Castiglione et al., 2003; Rasouli, 2016; Copperman et al., 2016). Most studies presented aggregate validation for different model kinds. For example, Bao et al. (2016) focused on two DTs only. Similarly, Copperman et al. (2016) described rail ridership. Bekhor et al. (2014) compared total vehicles kilometers travelled (VKT).

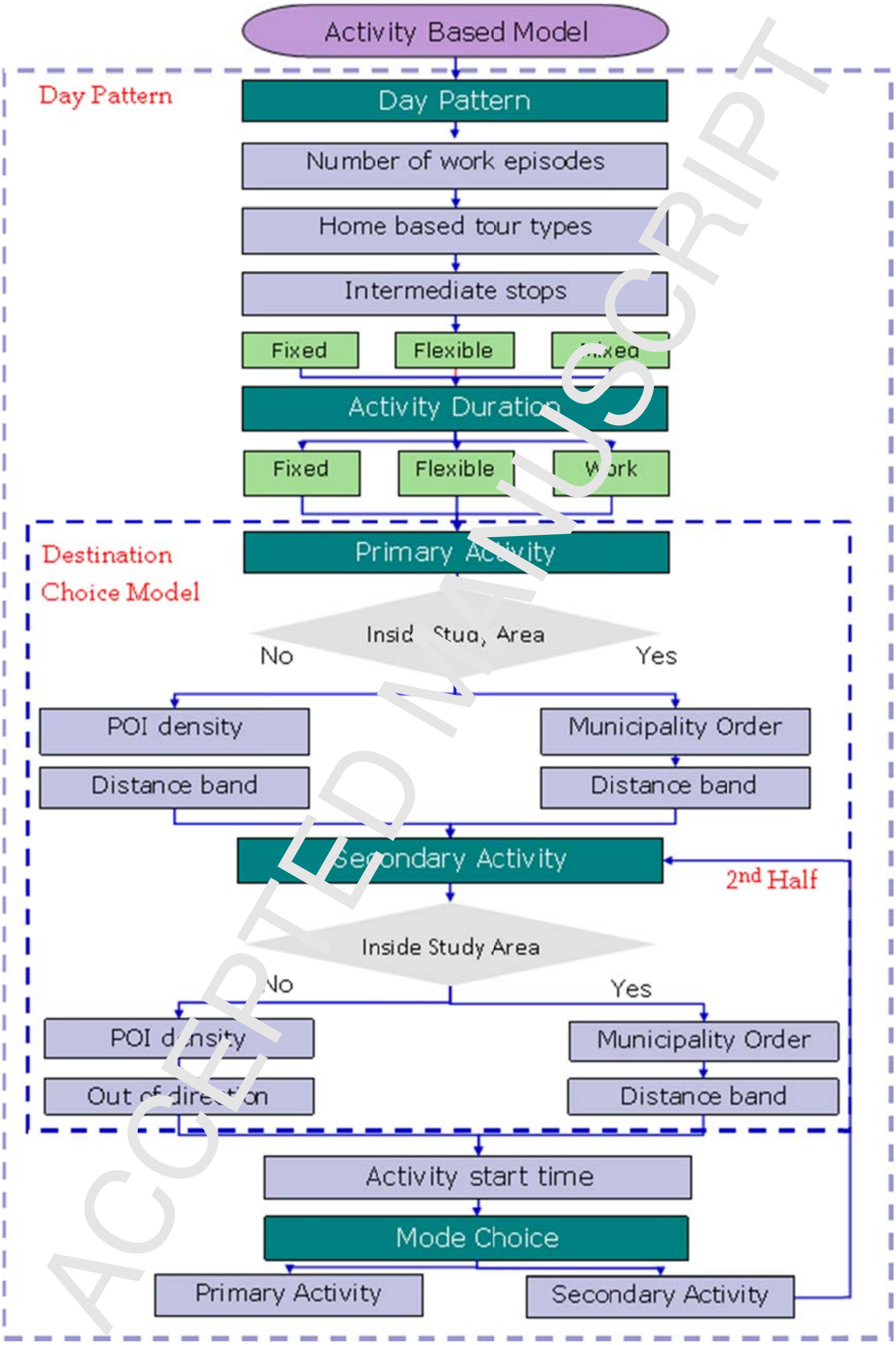
There is also a study that only described a generic validation framework for ABMs (Prelipcean et al., 2015). Drchal et al. (2016) described a Validation Framework for Activity-based Models (VALFRAM). The authors compared two basic system properties i.e. activities and trips across time, space and the structure (i.e. activity count and the travel mode used across activities). The study validated the model results using real-world activity-travel diary data and found a close relationship between both. Petrik et al.,(2018) discussed a variety of measures to compare the results of the two different model runs of an ABM in different settings to analyze model outcome uncertainty. They compare counts of tours, trips and stops for each activity, mode, location and a combination of them. The validation studies also vary with respect to the level of aggregation. For instance, Veldhuisen et al. (2000) compared origin-destination matrices at regional level. Furthermore, few studies also included socioeconomic attributes and described stratified model validation per population segment. Cools (2011) measured distance traveled across age and gender groups. Rasouli (2016) measured and presented validation results according to gender at the level of TAZs and study area. Besides these, Castiglione (2003) also included vehicle ownership in the validation criteria.

Literature suggests that the variation increases as the level of disaggregation increases. Therefore, it is important to assess model validation against individual attributes such as age, gender, vehicle ownership etc. Furthermore, rather than simply comparing the count, the emphasis should be on the distributions of activities and trips in time and space. Another important aspect for ABM validation is data availability. Since, an ABM not only needs to be validated for trips but also for activities, therefore, only traffic count data shall not suffice.

The above discussion emphasizes that it is essential to check the consistency of the model outputs when an ABM framework is modified before any transport related policies are tested. Additionally, to the best of our knowledge, there exists no study that integrates residents' external trips within the ABM and presents its validation. This study aims to address these gaps. The validation measures proposed in this study can also be used for validating other extensions in the ABM.

## 3. Research Framework

This section describes a framework to model residents' external travel as part of the complete activity-travel schedule in FEATHERS which is operational for Flanders, Belgium. A detailed functioning of FEATHERS is described in Bellemans et al. (2010), therefore, this paper only focuses on the components that are developed or modified to include the resident external travel within the current framework (Figure 1). These modifications include defining a CA, modifying destination choice models and the use of the open-source land use data in the destination choice models. Within the activity pattern model, first, the number of work episodes are determined followed by the generation of home-based tours. Then, for each tour, intermediate activities are determined along with their placement i.e. before or after the tour's primary activity. The intermediate activities are categorized as fixed [bring get, other] or flexible [shopping, services, social, leisure and touring]. Once each of the activity in the schedule is determined then their duration is modeled. The duration is categorized into three categories: short, medium and long. These categories have different time ranges as per the activity type. For example, a *medium* shopping activity may have lesser duration than a *short* leisure activity. For location choice, the first decision is the activity destination inside or outside the study area. Based on this decision, relevant Decision Trees (DTs) are triggered to estimate accurate location at the subzone level. The last step before the mode choice is the activity start time hour. At this moment only the hour is determined when the activities will take place, exact timings are randomly chosen within the 1-hour periods and are only available once all of the decisions have been made. The last decision is related to the transport mode for each activity. For each following DT, the schedule decisions simulated earlier are also included in the explanatory variables. The pseudo code of FEATHERS framework is shown in Figure 2.




194 *Figure 1: Framework to incorporate External activity-travel in Activity-Based Model FEATHERS*

```
Input: household travel survey, open-source land use, network and socioeconomic data, zonal information of
study area and the Catchment Area
For each individual do
    Initiate: person daily activity schedule
    Predict: daily activity pattern
    For each activity pattern do
        Initiate: related explanatory variables
        Predict: number of work episodes
        Predict: home based tour type HB, HBWS, HBWI1, HBWI2, HBWI12 or HBO
        For each home based work tour 1st half do
            Initiate: tour related variables
            Predict: intermediate activity type
        End
        For each home based work tour 2nd half do
            Initiate: tour related variables
            Predict: intermediate activity type
        End
        For each home based other tour do
            Initiate: tour related variables
            Predict: intermediate activity type fixed flexible mixed
            If intermediate activity type is fixed then
                Initiate: tour related variables
                Predict: intermediate stop activity fixed type
            Else If intermediate activity type is flexible then
                Initiate: tour related variables
                Predict: intermediate stop activity flexible
            Else
                Initiate: tour related variables
                Predict: intermediate stop activity fixed / flexible
            End
            For each activity do
                Initiate: explanatory variables
                Predict: activity duration
                For each primary episode do
                    Predict: destination in study area or catchment area
                    If destination is in study area then
                        Initiate: detailed land use explanatory variables
                        Predict: destination using municipality order and distance band
                        Predict: activity start time
                        Predict: transport mode
                    Else
                        Initiate: open source land use explanatory variables
                        Predict: destination using POI density and distance band
                        Predict: activity start time
                        Predict: transport mode
                    End
                    For each secondary episode do
                        Predict: destination in study area or catchment area
                        If destination is in study area then
                            Initiate: detailed land use explanatory variables
                            Predict: destination using municipality order and distance band
                            Predict: activity start time
                            Predict: transport mode
                        Else
                            Initiate: open source land use explanatory variables
                            Predict: destination using POI density and out-of-direction distance
                            Predict: activity start time
                        Predict: transport mode
                    End
                End
            End
        End
    End
End
output: results statistics for verification
output: daily activity schedules
```

*Figure 2: Pseudo code of FEATHERS simulation framework*

## 3.2 Defining Catchment Area

The primary region of interest for which an ABM is to be developed is defined as the study area. The external region adjacent to the study area is defined as the CA. The spatial unit of the CA should be the same as of the study area to avoid inconsistencies in the models. The spatial units are defined in FEATHERS at three levels: superzones (municipalities), zones (city) and subzones (TAZs). Depending on the size, a municipality may contain more than one city and a city may contain more than one TAZs.

In the proposed approach, the first step is to define the study area as per the modeling needs and collect the HTS data from a sample population within the boundary of the study area. Then, based on the travel pattern of the individuals in the HTS, a CA is defined. *The CA should be demarcated around the study area in a way such that it includes the farthest location that is used to perform an activity.*

This goes without saying that few *outliers* such as exceptionally long-distance trips should be excluded before defining the CA. This exemption is observed because of various reasons. First, the number of trips decreases as the distance from the study area increases which makes the model development cumbersome with the limited observations. Second, the probability that the individuals performing such trips will return back to their home within the simulated time period (typically 24 hours) is very less. Therefore these trips should be modeled as long-distance trips through the framework defined by Baqueri et al. (2018). Third, in case of an international border in the CA, there are also other issues such as the inaccessibility to TAZs specifications and dissimilarity in land use data which may generate unwanted model complexities.

For example, consider developing an ABM for Mechelen, a city in Flanders (Dutch speaking part of Belgium) with Brussels and Antwerp in its vicinity. Based on the OVG - household travel survey data of Flanders (Janssens et al., 2014), around 30% of the individuals travel outside Mechelen while the majority of the activities are performed within Flanders. Furthermore, only 1.4% of individuals commute to Wallonia (French-speaking part of Belgium) from Flanders due to the language barrier (Horckmans, 2017), which is quite low to train and test the model. Therefore, an ABM for Mechelen Flanders in included in the CA while Wallonia is discarded.

## 3.3 Destination choice model

The destination choice models in FEATHERS are built using DT with a multi-level decision hierarchy to specify the location of an activity. The first DT shortlists locations on the basis of predicted *Municipality Order* class. The municipality order is defined on the basis of attractiveness of a location and its distance from individual's current location. It is currently categorized in four categories, however, it can also be taken into continuous form when required. The second DT further narrow down locations on the basis of *Distance Band* (DB). The DB categorizes locations into classes on the basis of circular distance from the current location of the individual. Finally a location is randomly chosen from the remaining shortlisted locations belonging to the specified class of municipality order and the DB.

This methodology is first applied to the primary activity i.e. the main activity of the tour and then applied to the secondary activities of the tour. However, all decisions related to the primary activity are made first and then incorporated into the DTs of the secondary activities as the primary activity decisions directly influence on secondary activities.

### 3.3.1 Top level models

It is imaginable that the detailed land-use information, which has been obtained for the study area, may not be available for the CA. This is largely subjected to the limited resources or even unavailability of the information such as in case the study area is defined at the country level. Therefore, two top-level models are introduced in the current framework (shown in the decision box in Figure 1) each for the primary and the secondary activities which intent to identify if the activity will take place in the SA or the CA. If the activity will take place in the SA then the detailed information is used, otherwise only the variables formulated from open source platforms are used in estimating sub-models. Land use characteristics such as type, opening time, area, and employment and transport network attributes such as travel time, transit availability, price, and frequency can be obtained from open source platforms for developing destination choice models, mode choice models and time-of-day models. Some examples of the relevant Open source platforms are OpenStreetMap (OSM) (OpenStreetMap contributors, 2017) and Google API (Google Developers, 2017)). This is the first decision for assigning locations to activities, therefore, it is referred as the top-level model.

Some may argue that the inclusion of the top-level models (to define *if the activity shall be conducted in the SA or the CA)* in the decision hierarchy process is against the intuition as the SA boundary is simply a modeling term. While, in reality, an individual may not even be aware of the study area boundary let alone its inclusion in the decision process. However, this claim may not be true as the boundary of the study area has a practical significance whether it represents an international, provincial or a state-wide border or even a city- jurisdiction because individuals *do* consider these boundaries before choosing a destination.

For example, a Dutch citizen considers crossing the boundary between Netherlands-Belgium and Netherlands-Germany to commute as an equivalent to traveling 35 and 46 extra minutes respectively (Pieters et al., 2012). This border-crossing resistance is, however, less for shopping activity because of the same currency across the border. Similarly, the top level model may also be relevant in case of inter-regional travel. For example, as mentioned above, on average only 1.4% of individuals commute to Wallonia from Flanders due to the language barrier (Herckmans, 2017). Likewise, the statewide travel demand models are widespread in the USA which validates the fact that the inter-state travel is not so common. Furthermore, this decision-making impression may also be valid for the ABMs that are developed at the metropolitan-level and the boundary holds a toll cordon e.g. as in Paris during weekdays.

### 3.3.2 New Decision Trees

The inclusion of a top-level model also affects other subsequent location choice decisions. Therefore, 15 DTs are developed/modified to accommodate for the modified decision-hierarchy process for destination choice.

Tour's main Activity is defined as primary activities in FEATHERS. The DT *Choose Primary Location in Study Area or Catchment Area* defines if the primary activity will be performed in the CA or not. The need for this DT is described in section 3.3.1. Depending on the location two more DTs are used to determine precise activity location, i.e. the TAZ where the activity shall be performed. For activities to be conducted inside the CA, the first DT is *Choose POI Density Catchment Area* that identifies the POI density class in which the activity shall be conducted. The second DT for determining location is *Choose Distance Band Catchment Area* that identifies the distance band in which the activity shall take place. The distance band and POI density here are discretized into five classes which can be modified as required. For activities that are to be taken place inside the study area, the same DTs are used as in the model without the CA.

Activities other than the tour's main activity are defined as secondary activities in FEATHERS. These are distinguished in the activity-skeleton according to their placement before or after the primary activity. The activities performed before the primary activity are considered as 1st half while others are considered as 2nd half. The DT *Choose Secondary Location In Study Area Or Catchment Area 1st half* determines if the secondary activity that is to be conducted before the primary activity within the same tour will take place in or outside the study area. This is the top-level model for secondary activities (defined in section 3.2.1).

For the activities to be taken place inside CA, the DT *Choose Secondary Location in Catchment Area 1st half* is activated. An important variable in the DT is the *out-of-direction* travel distance which indicates that extent to which an individual deviates from a *straight line* between home and the primary activity location (equation 1). Similar DTs are used for determining locations of secondary activities that are to be performed after the primary activities.

$$Out - of - direction\ distance = [distance_{H\ to\ SL} + distance_{SL\ to\ PL} - [distance_{H\ to\ PL}] \quad (1)$$

Where H = home location, SL= secondary location and PL = primary location

The DTs for CA solely rely on individual's socioeconomic attributes, land use information obtained from open-source platforms, and already simulated activity-travel decisions from the higher order models but they do not incorporate any detailed land use and network information as it may not be available for the CA.

### 3.4 Relationship between open source and detailed land use information

Since the open-source land use information is incorporated in the DTs, therefore, it is important to verify its quality. This can be checked by comparing the open source land use information with the detailed land use information available for the study area. Figure 3 compares the land use information of Flanders, Belgium obtained from the official data source (Statbel, 2017) with the data obtained from the OpenStreetMap. The results show a strong association between commercial land use area from the official data source and the Points of Interest (POI) data from OpenStreetMap (OSM) in each Traffic Analysis Zone (TAZ). Furthermore, besides commercial land use, few other land use types also have a strong correlation with the POIs such as buildup and the transport land area (Table 1). This association (between official and open source land use data) may differ from region to region, but we believe a similar level of consistency of open source data, so our modeling methodology can be valid.

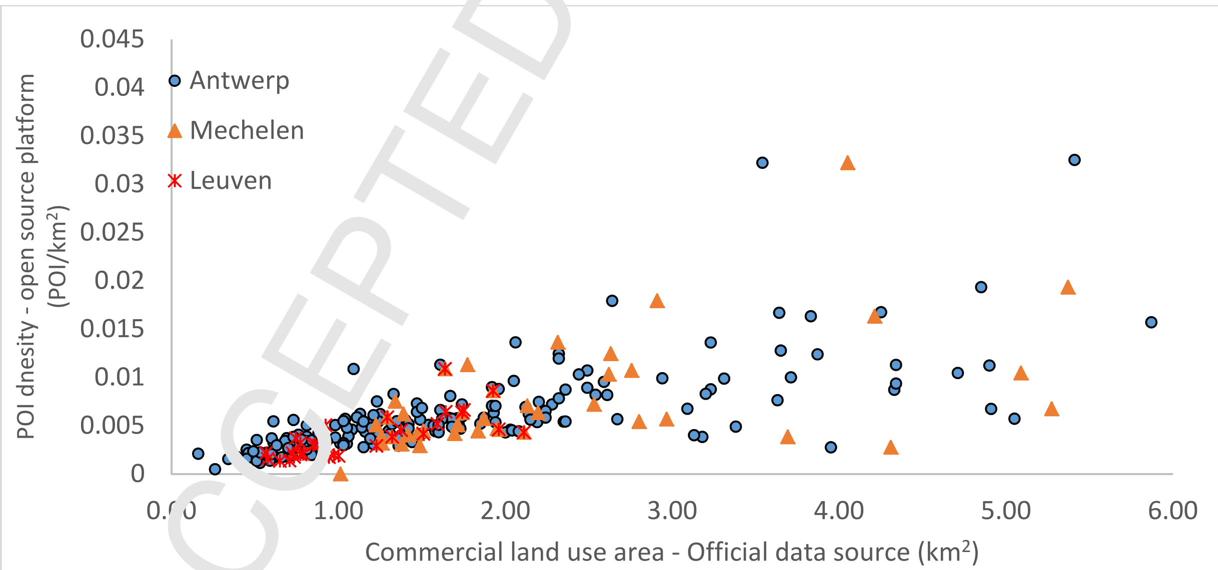



*Figure 3: Relationship between open source and official land-use data*

*Table 1: Correlation with official land use data and POI obtained from the open source platform*

| Land use type ($km^2$) | Correlation with POIs (number) |
|---|---|
| Commercial | **0.84** |
| Buildup land | **0.54** |
| Transport land | **0.51** |
| Public | 0.47 |
| Residential | 0.40 |
| Recreation Open area | 0.34 |

Highly correlated variables are marked in bold

# 4. Case study

This section describes the application of the above proposed FEATHERS framework on three study areas and the results obtained.

## 4.1 Implementation study areas

Currently, FEATHERS is operational for Flanders, Belgium and to test and validate the proposed framework, smaller regions in Flanders are defined as the study areas (Figure 4). These study areas have the following properties:

- Are medium-sized regions with a population between 0.5 to 1 million and an area around 1,000$km^2$
- Population density varies between 400persons/$km^2$ to 1,000persons/$km^2$.
- Around 25 - 35% of the residents perform external travel (obtained from BELDAM data (Hollaert et al., 2012)).
- Are a major trip attractor themselves and/or surrounded with a major trip attractor in their vicinity that influence external travel.

The details and the significance of these regions to test the proposed methodology are further defined.

### 4.1.1 Antwerp region

Antwerp region is located in the north of Flanders. It is the most populated province in Belgium with a population of 1.8 million. It is an attractive region with a port that generates a lot of commercial activity. Approximately 30% of the individuals tend to perform their activities outside the region, therefore, it shall be useful to check the distribution of activity types, and in particular work activities, in and outside the region.

### 4.1.2 Mechelen region

Mechelen is a *home city* for a lot of individuals who work in Brussels. Besides, Mechelen is equally distant between Brussels and Antwerp which makes it an interesting case to evaluate the proposed methodology. In order to define a relevant study area, a 20km radius around Mechelen city is considered having a population of around 0.5 million. Approximately 34% of the residents perform external travel.

### 4.1.3 Leuven region

Leuven is located in Southern part of Flanders. It is surrounded by Brussels in its East which is an attractive region and attracts a lot of external travel. Therefore, it shall be interesting to implement this framework in Leuven region. The population of Leuven region is approximately 0.5 million and nearly 30% of the residents perform external travel.

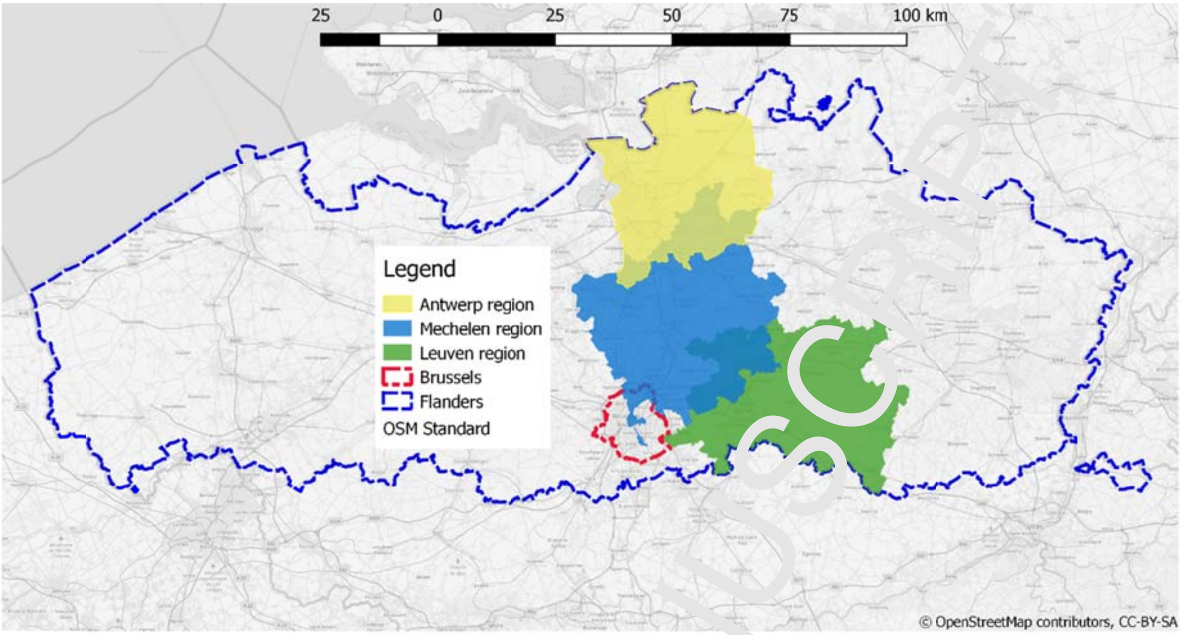



*Figure 4: Study Areas Boundaries of Antwerp, Mechelen and Leuven region*

## 4.2 Results

The results obtained from running FEATHERS on these regions are described in this subsection. Only the individuals belonging to these study areas are used for model training and developing the synthetic population.

Table 2 compares the actual and predicted trips, tours, mode split and distribution of activities in the study area and Catchment Area. On aggregate, a close relationship is found between FEATHERS for full Flanders and for smaller study areas. For instance, earlier 23% of individuals performed an activity outside Antwerp region and in FEATHERS with CA setup 24% performed at least one activity in CA. Similar results are obtained for Mechelen (34%; 31%). However, some differences are present in Leuven (29%; 36%). However, there are some differences in mode split in Antwerp region where a larger share of trips have been assigned to cars against public transport users.

*Table 2: Aggregate results with and without Catchment Areas*

| Parameter | Antwerp | | Leuven | | Mechelen | |
|---|---|---|---|---|---|---|
| Peak Activity Start Time | Without CA | With CA | Without CA | With CA | Without CA | With CA |
| Average Time spent travelling (min) | 44.31 | 46.74 | 57.48 | 60.79 | 55.65 | 52.43 |
| % of trips in Peak hour | 8.59 | 9.93 | 8.90 | 9.34 | 8.9 | 8.5 |
| Work Activity % | 23.54 | 23.02 | 26.54 | 24.91 | 26.54 | 25.42 |
| Education Activity % | 15.99 | 13.47 | 21.91 | 15.91 | 21.92 | 12.94 |
| Daily Shopping Activity % | 21.51 | 20.60 | 15.28 | 21.99 | 15.28 | 19.35 |
| Non-Daily Shopping Activity % | 8.28 | 10.10 | 7.46 | 8.77 | 7.46 | 8.27 |
| Services Activity % | 10.15 | 12.06 | 9.12 | 8.65 | 9.12 | 10.81 |
| Car % | 42.23 | 46.8 | 48.93 | 47.52 | 48.94 | 46.49 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Public Transport % | 10.36 | 9.16 | 29.68 | 10.64 | 7.6 | 11.84 |
| Non-Motorized transport % | 31.07 | 24.17 | 7.59 | 23.23 | 2[illegible].68 | 24.06 |
| Simple tour % | 62.35 | 63.68 | 60.58 | 60.60 | 64.34 | 20.40 |
| 2-activity tours % | 21.73 | 21.26 | 22.61 | 22.49 | [illegible]0.40 | 9.24 |
| % of work Activities in CA | 33.96 | 30.61 | 45.08 | 44.6 | 35.85 | 38.06 |
| % on individuals travelling to CA | 22.6 | 23.8 | 28.9 | 35.8 | 33.7 | 30.1 |

Table 3 shows the improvement in the contingency matrix of DTs after the proposed changes; inclusion of a top-level model and POI density in the DTs. The DTs determine various aspects of the activity-travel pattern such as *activity start time, duration, destination choice, intermediate stop type* etc. It can be observed that these changes and in particular POI density considerably increased the DT's explanatory power in many cases. These improvements account even above 60%. An exception, in this case, is for DT choose *Number of Work Episodes* where the overall model explanatory power is reduced. However, it should be noted that the model accuracy is still above 75% in each region, therefore, these are negligible reductions.

POI density is found significant in new DTs created to specifically model location choice of primary activity. However, it is found significant in only one DT for secondary activity. The results are further elaborated in Discussion (section 6).

*Table 3: Improvement in Decision Trees in Activity-Based Model for medium-sized study area as compared to the Full-scale model*

| Decision Tree / Study area | Antwerp | Mechelen | Leuven |
|---|---|---|---|
| Choose Number Of Work Episodes | -1.49* | -1.55* | -1.27* |
| Choose Home-Based Tour Types Sequence | 5.42* | 45.70* | 26.65* |
| Choose HBWI1 Intermediate Stop Activities | 37.84 | 41.33 | 22.28 |
| Choose HBWI2 Intermediate Stop Activities | -0.04 | 1.61 | 27.58 |
| Choose HBWI12 Intermediate Stop Activities | 56.83 | 22.98 | 39.33 |
| Choose HBO Intermediate Stop Types Fixed Flexible Mixed | 1.34 | 2.64* | -2.83* |
| Choose HBO Intermediate Stop Activities Fixed | 2.80 | -1.20 | 2.31 |
| Choose HBO Intermediate Stop Activities Flexible | 1.97 | 3.05* | 0.87* |
| Choose HBO Intermediate Stop Activities Mixed | 8.31 | 5.86* | 16.16* |
| Choose Duration First Work Activity | -3.61 | -1.86 | -1.94 |
| Choose Duration Second Work Activity | 7.31 | 4.33 | 13.49 |
| Choose Duration Fixed Activities | 1.99 | 2.27* | 0.12* |
| Choose Duration Flexible Activities | 14.79* | 13.69 | 19.56 |
| *Choose Primary Location In Study Area Or Catchment Area* | x | x | x |
| *Choose Primary Location In Home Municipality* | | | x |
| *Choose Primary Location In Home Subzone* | x | x | x |
| *Choose Order Municipality* | | | |
| *Choose Nearest Order Municipality* | | | |
| *Choose Distance Band Superzone* | | | |
| *Choose POI Density Superzone Catchment Area* | x | x | |
| Choose Start Time Hour of Home Based Tour Primary Episode | 2.25 | 3.69 | 4.92 |
| *Choose Transport Mode Primary Episode* | 59.86 | 57.66 | 62.11 |
| +*Choose Secondary Location In Study Area Or Catchment Area 1st half* | | | |

| | | | |
|---|---|---|---|
| *+ Choose Secondary Location Type In Study Area 1st half* + | | | |
| *+ Choose Secondary Location In Study Area 1st half* + | | | |
| *+ Choose Secondary Location In Catchment Area 1st half* + | | | |
| Choose Start Time Hour of Home Based 1st Half Tour Secondary Episode | 5.37 | 1[illegible].4[illegible]* | 3.25* |
| Choose Transport Mode Secondary Episode 1st half tour | -3.07 | 3.[illegible] | -9.12 |
| *+Choose Secondary Location In Study Area Or Catchment Area 2nd half* + | x | | |
| *+ Choose Secondary Location Type In Study Area 2nd half* + | | | |
| *+ Choose Secondary Location In Study Area 2nd half* + | | | |
| *+ Choose Secondary Location In Catchment Area 2nd half* + | | | |
| Choose Start Time Hour of Home Based 2nd Half Tour Secondary Episode | 0.[illegible]6 | 1.92 | -5.63 |
| Choose Transport Mode Secondary Episode 2nd half tour | -[illegible].80 | 3.08 | -3.55 |
| Choose Start Time Hour of Home Based Tour Last Home Episode | 3.1[illegible] | 4.78 | 4.58 |
| Choose Transport Mode of Home Based Tour Last Home Episode | [illegible].27 | 2.57 | -1.40 |

* sign shows DTs in which POI density is found to be significant, + sign indicates new DTs created to specifically model external travel, x= DTs where POI density is found to be significant, HBW= Home based Work, HBO=home based other, I1 = secondary activity before the primary activity, I2 = secondary activity after the primary activity

# 5 Model Validation

The proposed framework-changes also stresses its accurate validation in order to evaluate its effectiveness and dependability. For instance, the top-level model may result in too many or too few individuals going to the CA. Similarly, there is a possibility that the activities in CA may result in larger time spent traveling or a substantial shift in the transport mode choice. Besides, the activity pattern may be altered that may substantially affect tours. Therefore, a validation framework for an ABM should validate activities, trips as well as tours.

Therefore, this section describes the statistical validation of the results obtained. First, a validation framework is defined followed by the description of the two models used for validation and lastly the validation metrics produced.

## 5.1 Validation Framework

The validation framework presented in this study extends the framework proposed in earlier studies (Drchal et al., 2016; Petrik et al., 2018) in three dimensions: (1) expands the scope of *structure* to model distribution of activities between SA and CA (2) includes the tour dimension in the validation besides activities and trips and (3) disaggregate validation of the proposed measures against socioeconomic attributes of the population. In total, 11 benchmarks are proposed to comprehensively validate ABM results (Table 4Table 4). These benchmarks complement the outcome of the DTs associated with the *activity pattern, start time, duration, location* choice and *mode choice*. These benchmarks are further described according to type.

*Activities:* Activities are the driving force behind the Activity-based Travel Demand Models (Ben-akiva et al., 1996). Therefore, it is important to carefully validate various aspects of activities. This paper describes eleven measures for validating activity distribution across space, time and structure (Table 4). An important remark here is that there is no concept of CA in the ABM developed for Flanders model, therefore, some post-processing is required before validation *Activity Distribution in CA and SA*. For this, the locations outside the study area in the medium-sized model are considered as CA in the output of the full-scale model. This process is repeated for each study area separately.

*Trips:* Three measures are suggested for comparing trips between a full-scale and a medium-sized ABM. These include the distribution of trips performed across travel modes and also the time spent traveling.

*Tours:* Tours are also a vital aspect of ABM as these link together the two major components of ABM i.e. activity and travel. Therefore, two measures are incorporated to validate the tour-consistency between predicted and actual data. These measures define the number of tours and their complexity.

## 5.2 Validation Model Description

The most important step to validate model results, after defining a validation framework, is the availability of a data source that is not used in the model development. In this study the model output of FEATHERS for Flanders region without the CA setup have been considered for validation. For validating, the outputs of the model without the CA are post-processed and the locations are labeled as inside study area or CA as in the model with the CA.

## 5.3 Aggregate Validation

Table 5 shows aggregate analysis of the proposed benchmarks in Antwerp, Leuven and Mechelen region. None of the benchmarks are found to be statistically different between both the models at 10% significance level in Antwerp while some differences are found in other regions.

*Table 4: Validation benchmarks of the Activity-Based Model*

| S. No | Benchmarks | Level | Assembly | Task |
|---|---|---|---|---|
| 1 | Time spent on each activity type | Activities | Time | Distribution of time spent on each activity type. Only out-of-home activities are considered |
| 2 | Activity start time | Activities | Time | Distribution of activity start time in 30-minute time bins. |
| 3 | Activity Distribution in CA and SA | Activities | Space | Distribution of share of each activity-type in total activities performed in CA |
| 4 | Types of activities performed* | Activities | Structure | Distribution of n different activities performed across m individuals. For ease, only out-of-home activities are considered. |
| 5 | Number of total activities | Activities | Structure | Distribution of total activities performed across individuals |
| 6 | Number of out-of-home activities | Activities | Structure | Distribution of number of out-of-home activities performed across individuals |
| 7 | Number of in-home activities | Activities/ Tour | Structure | The number of times an individual returns home within a simulated day. |
| 8 | Tour complexity | Tour | Structure | Distribution of share of *a* activities performed by *m* individuals before returning home |
| 9 | Trips by each mode | Trips | Structure | Distribution of percentage of trips by each travel mode |
| 10 | Types of transport mode use | Trips | Structure | Distribution of *i* transport modes used in trips by *m* individuals |
| 11 | Time spent traveling | Trips | Time | Distribution of time spent traveling in 10-minute bins |

* FEATHERS distinguishes out-of-home activities in 10 categories: Work, Bring/get, Shopping (daily), Shopping (non-daily), Services, Social visits, Leisure, Touring and Other.

*Table 5: Aggregate validation of proposed benchmarks using Kolmogorov-Smirnov test*

| | Antwerp Region | Mechelen Region | Leuven Region |
|---|---|---|---|
| Criteria | P-Value | P-Value | P-Value |
| Percentage of trips by each mode | 1.00 | 0.97 | 1.00 |
| Types of transport mode use | 1.00 | 1.00 | 1.00 |
| Time spent travelling | 0.70 | 1.00 | 0.40 |
| Types of activities performed | 0.99 | 0.7[illegible] | 0.98 |
| Number of in-home activities | 0.98 | 0.98 | 1.00 |
| Number of out-of-home activities | 1.00 | 1.00 | 1.00 |
| Number of total activities | 1.00 | 1.00 | 1.00 |
| % Of time spend on each activity | 1.00 | 1.00 | 0.98 |
| Tour complexity | 1.00 | 0.66 | 1.00 |
| Activity start time | 1.00 | 0.87 | 0.79 |
| Activity Distribution in CA and study area | 0.98 | 0.63 | 0.63 |

## 5.4 Disaggregate Validation

This section describes disaggregate analysis of the proposed benchmarks. Five socioeconomic characteristics (age, work status, driving license, income, and number of cars) are chosen for disaggregate analysis (Table 6). Amongst these, the first three represent individual characteristics while the latter two signify household attributes. The disaggregate validation of each of these criteria is further described for each study area separately.

*Table 6: Classes of socioeconomic variables*

| Group | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Age (years) | 18-34 | 35-54 | 55-64 | 65-74 | 74+ |
| Work Status | Unemployed | Employed | - | - | - |
| Driving License | No | Yes | - | - | - |
| Socioeconomic Class [Income (€)] | 0-1249 | 1250-2249 | 2250-3249 | 3250+ | - |
| Number of Cars | 0 | 1 | 2 or more | - | - |

Some differences are found in the benchmarks in each region (Table 7-9). For instance, the *distribution of Activities in CA* is found to be significantly different between age group four (65-74 years) and also in case of Socioeconomic Class (*SEC)* group one. In total, three distributions are found to be different in Mechelen and it is observed that these classes have lesser observations than average. Table 9 shows validation results for Leuven region. *Time spent on activities* is significantly different for age group five (75 years or above). Similarly, *time spent traveling* is also found to be significantly different for households having no car. This may be due to the fact that unlike most of the other measures, time spent on activities is arbitrarily grouped using 10-minute intervals. The result changes if another value is used for defining the significance level.

*Table 7: Disaggregate results of Kolmogorov-test for Antwerp region*

| Criteria / Class | Age | | | | | Work Status | | License | | Socioec | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Criteria / Class | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 1 | 2 | 1 | |
| Activity Start Time | 1.00 | 0.79 | 0.79 | 1.00 | 0.97 | 0.79 | 0.97 | 0.79 | 0.98 | 0.97 | 0 |
| Share of each transport Mode | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.77 | 1.00 | 0.77 | 1 |
| Number of modes used by each individual | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.70 | 0.70 | 0.70 | 0.70 | 0.70 | 0 |
| Time spent travelling | 0.40 | 0.99 | 0.99 | 0.76 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0 |
| Types of activities performed | 0.96 | 1.00 | 0.27 | 0.98 | 1.00 | 0.98 | 1.00 | 0.63 | 0.66 | 0.63 | 0 |
| Number of In-home activities | 0.93 | 0.66 | 0.93 | 0.66 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1 |
| Number of out-of-home activities | 1.00 | 1.00 | 0.08* | 0.93 | 0.93 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1 |
| Number of total activities | 0.93 | 0.93 | 0.93 | 0.66 | 0.93 | 1.00 | 1.00 | 0.93 | 1.00 | 1.00 | 1 |
| Time spent on activities | 0.66 | 0.98 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 | 0.98 | 1.00 | 0.98 | 1 |
| Tour Complexity | 0.87 | 1.00 | 1.00 | 0.82 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.82 | 0 |
| Distribution of Activities in CA | 0.63 | 1.00 | 0.63 | 0.96 | 0.52 | 0.63 | 0.63 | 0.63 | 0.63 | 0.96 | 0 |

*significantly different at 10% significance level

*Table 8: Disaggregate results of Kolmogorov-test for Mechelen region*

| Criteria / Class | Age | | | | | Work Status | | License | | Socioe |
|---|---|---|---|---|---|---|---|---|---|---|
| Criteria / Class | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 1 | 2 | 1 |
| Activity Start Time | 0.97 | 0.97 | [illegible] | 0.30 | 0.79 | 0.97 | 0.79 | 0.53 | 1.00 | 0.53 |
| Share of each transport Mode | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Number of modes used by each individual | 1.00 | 1.00 | 1.00 | 0.70 | 0.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Time spent travelling | 0.40 | 0.40 | 0.40 | 0.76 | 1.00 | 0.40 | 0.40 | 0.76 | 0.99 | 0.99 |
| Types of activities performed | 0.63 | 0.98 | 0.98 | 0.63 | 0.96 | 0.66 | 0.98 | 0.96 | 0.98 | 0.96 |
| Number of In-home activities | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Number of out-of-home activities | 1.00 | 0.93 | 1.00 | 0.66 | 0.93 | 1.00 | 1.00 | 1.00 | 0.93 | 1.00 |
| Number of total activities | 0.93 | 0.93 | 1.00 | 0.38 | 0.18 | 1.00 | 0.93 | 0.93 | 0.93 | 1.00 |
| Time spent on activities | 0.66 | 0.66 | 0.66 | 0.08* | 0.27 | 0.66 | 0.28 | 0.96 | 0.28 | 0.27 |
| Tour Complexity | 0.82 | 1.00 | 0.82 | 0.82 | 0.82 | 0.82 | 1.00 | 0.87 | 0.87 | 0.82 |
| Distribution of Activities in CA | 0.63 | 0.96 | 0.27 | 0.02* | 0.63 | 0.66 | 0.27 | 0.96 | 0.96 | 0.09* |

*Table 9: Disaggregate results of Kolmogorov-test for Leuven region*

| Criteria / Class | Age | | | | | Work Status | | License | | Socioe |
|---|---|---|---|---|---|---|---|---|---|---|
| Criteria / Class | 1 | 2 | 3 | 4 | 5 | 1 | 2 | 1 | 2 | 1 |
| Activity Start Time | 0.49 | 0.96 | 0.96 | 0.77 | 0.49 | 0.30 | 0.79 | 0.53 | 0.53 | 0.07* |
| Share of each transport Mode | 1.00 | 1.00 | 1.00 | 0.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.70 |
| Number of modes used by each individual | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Time spent travelling | 0.76 | 0.76 | 0.16 | 0.20 | 0.40 | 0.16 | 0.40 | 0.40 | 0.40 | 0.76 |
| Types of activities performed | 1.00 | 0.66 | 0.63 | 0.27 | 0.96 | 1.00 | 0.98 | 0.63 | 0.66 | 0.27 |
| Number of In-home activities | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.93 | 1.00 | 1.00 |
| Number of out-of-home activities | 0.66 | 1.00 | 1.00 | 0.93 | 1.00 | 1.00 | 1.00 | 0.93 | 1.00 | 1.00 |
| Number of total activities | 0.66 | 1.00 | 1.00 | 0.66 | 0.66 | 0.93 | 0.93 | 0.38 | 1.00 | 1.00 |
| Time spent on activities | 0.66 | 0.98 | 0.98 | 0.63 | 0.09* | 0.28 | 0.98 | 0.63 | 0.98 | 0.63 |
| Tour Complexity | 0.33 | 0.87 | 1.00 | 0.33 | 1.00 | 1.00 | 1.00 | 1.00 | 0.87 | 0.82 |
| Distribution of Activities in CA | 0.63 | 0.96 | 0.96 | 0.09* | 0.09* | 0.63 | 0.63 | 0.89 | 0.27 | 1.00 |

# 6 Discussion

This paper describes a scheme to model residents external activity and travel by defining only the region of interest as the study area and its surrounding region as the CA. Defining the CA allows to model external activity-travel as part of complete schedule rather than modeling them separately through external travel models. Thus, the presented methodology allows to develop an ABM for a medium-sized region by addressing the issue of external travel. Furthermore, it also reduces data collection, model development and computational efforts as the HTS and synthetic population is only required for the study area. However, defining a medium-sized region as a study area also increases non-resident external trips in the study area. therefore, proper estimation of non-residents external trips is required in order to correctly calibrate the ABM. To address this issue, a comprehensive methodology is described to estimate non-residents external trips which only rely on the open-source platforms and the HTS. For details, the readers may refer to (Baqueri, Adnan & Bellemans, 2018; Baqueri, Adnan, Knapen, et al., 2018). Therefore, defining a medium-sized study area and properly estimating external trips is a better approach in terms of data collection and model development efforts for ABMs while estimating external trips through a non-data intensive approach.

The ABM framework proposed in this study has a generic skeleton and can be applied to any other ABM. An added value of this approach is the ability to test policy scenarios. For instance, What shall be the effect on residents' travel pattern of an improved transit service in the CA? or the effect of land use change in the CA on the distribution of activities within and outside the study area? Or implications of congestion charging around the boundary of the study area on total vehicle kilometers traveled?

There are some observations that require further explanation. For instance, the variable POI density is not found significant in the DTs that determine the location of secondary activity, except in one occasion. One reason behind this may be that the POI density is defined irrespective of the activity type that can be performed there. However, most open-source platforms allow categorizing POI according to the activity type such as work, education, shopping, etc. Thus, the POI densities can be calculated discretely for each activity type. This adaptation shall further enrich the DTs for each type of the secondary activities. Furthermore, the variation in the land use can also be effectively utilized by developing numerous indexes from the open-source data. Case in point is the Entropy Index measure which solely relies on the POI count and describes the land use as mixed or suitable only for a particular activity type (Baqueri, Adnan & Bellemans, 2018).

Another important aspect here to consider is the quality of the open-source data. For example, the correlation between the builtup area and POI density in Antwerp, Mechelen, and Leuven is 0.68, 0.67 and 0.85 respectively. This strong association between the two data sources improved the model explanatory power and especially the top-level model. The results may be different if the two data sources do not match with each other. Therefore, a successful implementation of the proposed approach heavily depends on the quality of the open-source data. Furthermore, the POI data represents the land use just as a point and does not distinguish them on the basis of area, height, and other attributes. Therefore, a multi-story land use could be considered equivalent to a single shop. For instance, the hospital in Leuven is a super entity where patients from all over Flanders visit, thus generating a lot of external travel. However, the lack of data on its area or other characteristics undervalues its prominence. This shall be a possible explanation behind differences in some validation measures in the Leuven region.

Besides, the availability of a land use (in terms of opening hours) is also relevant for assigning locations, which many open-source platforms either do not contain at all or allow its restricted usage. However, with the advancements in the Internet of Things (IoT), further detailed information can be obtained and utilized as per the availability and the modeling requirements. Few recent studies have described the potential usefulness of the open-source and the social media data for modeling travel behavior. For a comprehensive overview of the challenges and available opportunities in this regard, the readers may refer to Rashidi et al. (2017).

# 7 Conclusion and Future Work

This paper presented a framework to develop an ABM for medium-sized regions by allowing for residents' external activity-travel. Earlier studies separately modeled residents' external travel (i.e. outside the scope of the ABM) which resulted in many drawbacks such as the distortions in travel patterns as activity-locations are assigned only within the study area. Therefore, for an ABM to be effective in replicating the actual environment, an expanded study area is required to minimize the external travel.

In the proposed framework, the external locations are included in the destination choice models in the form of a CA as possible locations to perform an activity. The destination choice models are then modified with top-level models that determine the destination for each activity in the study area or CA. For activities to be performed inside the CA, a series of DTs are activated that collectively decide the destination. These DTs solely rely on individual's socioeconomic attributes, available activity-travel decisions, and open-source land use information but they do not require any detailed land use or network information as that may not be available for the CA. These modifications allow modeling external activity-travel as part of the daily travel pattern rather than estimating them through separate models which are not sensitive to policy measures. Furthermore, the proposed approach also provides an added flexibility to define the study area as per the modeling needs. These changes are implemented in ABM-FEATHERS and tested on three medium-sized regions in Flanders, Belgium. The results confirm clear advantages of the proposed methodology in terms of the decision hierarchy, model development, run-time and also data collection efforts if the ABM needs to be developed from scratch. Slight differences in validation are also found in one region where the POI density is not in a close relationship with the detailed land use data. This suggests that the availability of adequate land use information holds a central position in the proposed framework.

Furthermore, a comprehensive validation framework is also suggested to compare the model outputs obtained by defining complete Flanders as the study area and these medium-sized regions as the study areas. The validation measures include a comparison between activities, trips, and tours in terms of time, space and the structure. Furthermore, disaggregate validation is also analyzed using five socioeconomic characteristics (age, work status, driving license, income, and number of cars). The results confirm a close resemblance between both the models which suggests that an ABM can be developed for small-scale regions, once the question of external travel is addressed. This paper, therefore, shall pave the way for practitioners in developing an ABM for a medium-sized region.

The future work shall focus on further testing the applicability of the proposed approach. For instance, numerous policy scenarios can be tested in the study area or the CA or a case study of new transport policies/ services etc. can be studied. This way the added value of the framework can be quantified better by comparing it against a benchmark such as the full-scale ABM. This shall ultimately, therefore, lead towards developing ABM for medium-sized regions.

# References


Adnan, M., Pereira, F.C., Azevedo, C.L., et al., (2016). Evaluating Disruption Management Strategies In Rail Transit Using SimMobility Mid-Term Simulator□: A Study Of Singapore MRT North-East

Line. *96th Annual Meeting, Transportation Research Board*.

Adnan, M., Pereira, F.C., Miguel, C., et al., (2016). SimMobility□: A Multi-Scale Integrated Agent-based Simulation Platform. *Transportation Research Board (TRB) 95th Annual Meeting*, pp.1–18.

Arentze, T.A. & Timmermans, H.J.P., (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological*, 38(7), pp.613–633.

Auld, J. & Mohammadian, A.K., (2012). Activity planning processes in the Agent-based Dynamic Activity Planning and Travel Scheduling (ADAPTS) model. *Transportation Research Part A: Policy and Practice*, 46(8), pp.1386–1403.

Bao, Q. et al., (2016). *Activity-based travel demand modeling framework feathers: Sensitivity analysis with decision trees*,

Bao, Q. et al., (2015). Investigating micro-simulation error in activity-based travel demand forecasting: a case study of the FEATHERS framework. *Transportation Planning and Technology*, 38(4), pp.425–441.

Baqueri, S.F.A., Adnan, M., Knapen, L., et al., (2018). Modeling Distribution of External–Internal Trips and Its Intra-region and Inter-region Transferability. *Arabian Journal for Science and Engineering*, accepted,. Available at: http://link.springer.com/10.1007/s13369-018-3482-x.

Baqueri, S.F.A., Adnan, M. & Bellemans, T., (2018). Modeling External Trips: Review of Past Studies and Directions for Way Forward. *Journal of Transportation Engineering, Part A: Systems*, 144(9). Available at: http://ascelibrary.org/doi/10.1061/JTEPBS.0000179.

Bekhor, S., Kheifits, L. & Sorani, M., (2014). Stability analysis of activity-based models: Case study of the Tel Aviv transportation model. *European Journal of Transport and Infrastructure Research*, 14(4), pp.311–331.

Bellemans, T. et al., (2010). Implementation Framework and Development Trajectory of FEATHERS Activity-Based Simulation Platform. *Transportation Research Record: Journal of the Transportation Research Board*, 2175(1), pp.111–119.

Ben-akiva, M., Bowman, J.L. & Gopinath, D., (1996). Travel demand model system for the information era. *Transportation*, 23, pp.241–266.

Bowman, J.L.. & Ben-Akiva, M., (1998). The Day Activity Schedule Approach to Travel Demand Analysis. *Transportation*.

Bowman, J.L. & Bradley, M., (2006). *Activity-Based Travel Forecasting Model for SACOG*,

Castiglione, J., Freedman, J. & Bradley, M., (2003). Systematic Investigation of Variability due to Random Simulation Error in an Activity-Based Microsimulation Forecasting Model. *Transportation Research Record*, 1831(1), pp.76–88.

Childress, S., Nichols, B. & Coe, S., (2015). Using an activity-based model to explore possilbe impacts of automated vehicles. *Transportation Research Board 94th Annual Meeting*.

Cools, M. et al., (2011). Assessment of the Effect of Micro-simulation Error on Key Travel Indices: Evidence from the Activity-Based Model FEATHERS. In *Transportation Research Board 90th Annual Meeting*.

Copperman, R. et al., (2016). Development of a Risk Analysis Methodology for Quantifying the Uncertainty of Travel Demand Forecasts. Available at: http://tfresource.org/images/a/a6/ITM16_Development_of_a_Risk_Analysis_Methodology_for_Qua ntifying_the_Uncertainty_of_Travel_Demand_Forecasts.pdf.

Department for Transport, (2017). *NTEM Planning Data Version 7.2*,

Drchal, J., Č Ertický, M. & Jakob, M., (2016). Data driven validation framework for multi-agent activity-based models. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. pp. 55–67.

Gärling, T., Kwan, M.P. & Golledge, R.G., (1994). Computational-process modelling of household activity scheduling. *Transportation Research Part B*, 28(5), pp.355–364.

Gibb, J. & Bowman, J.L., (2007). Convergence of an activity-based travel model system to equilibrium: Experimental design and findings. In *Proceedings of the 11th National Transportation Planning and Application Conference of the Transportation Research Board*. Available at: http://www.trbappcon.org/2007conf/files/049 Gibb final.pdf.

Google Developers, (2017). Developers Site. Available at: https://developers.google.com/maps/documentation/business/places/ [Accessed November 3, 2016].

Hollaert, L. et al., (2012). *La mobilité en Belgique en 2010 : résultats de l'enquête BELDAM*,

Horckmans, M., (2017). Waalse werknemer gaat sneller taalgrens over dan Vlaamse collega. *Express Business*. Available at: https://nl.express.live/2017/01/26/waalse-werknemer-gaat-sneller-taalgrens-dan-vlaamse-collega/ [Accessed February 27, 2018].

Hussain, I. et al., (2016). Negotiation and Coordination in Carpooling. *Transportation Research Record: Journal of the Transportation Research Board*, 2542, pp.92–101. Available at: http://trrjournalonline.trb.org/doi/10.3141/2542-11.

Janssens, D., Declercq, K. & Wets, G., (2014). *Onderzoek Verplaatsingsgedrag Vlaanderen 4.5 (2012-2013)*,

de Jong, G. et al., (2007). Uncertainty in traffic forecasts: Literature review and new results for The Netherlands. *Transportation*, 34(4), pp.375–395.

Kaddoura, I., Kröger, L. & Nagel, K., (2017). An activity-based and dynamic approach to calculate road traffic noise damages. *Transportation Research Part D: Transport and Environment*, 54, pp.335–347.

Knapen, L. et al., (2012). Activity-Based Modeling to Predict Spatial and Temporal Power Demand of Electric Vehicles in Flanders, Belgium. *Transportation Research Record: Journal of the Transportation Research Board*, 2287, pp.146–154. Available at: http://trrjournalonline.trb.org/doi/10.3141/2287-18.

Lefebvre, W. et al., (2013). Presentation and evaluation of an integrated model chain to respond to traffic- and health-related policy questions. *Environmental Modelling and Software*, 40, pp.160–170.

Miller, E. & Roorda, M., (2003). Prototype Model of Household Activity-Travel Scheduling. *Transportation Research Record: Journal of the Transportation Research Board*, 1831, pp.114–121. Available at: http://trrjournalonline.trb.org/doi/10.3141/1831-13.

OpenStreetMap contributors, (2017). https://planet.osm.org.

Petrik, O. et al., (2018). Uncertainty analysis of an activity-based microsimulation model for Singapore. *Future Generation Computer Systems*. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167739X1830150X.

Pieters, M., de Jong, G. & van der Hoorn, T., (2012). Cross-border car traffic in dutch mobility Models. *European Journal of Transport and Infrastructure Research*, 12(2), pp.167–177.

Prelipcean, A.C., Gidófalvi, G. & Susilo, Y.O., (2015). Comparative framework for activity-travel diary collection systems. In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2015*. pp. 251–258.

Rashidi, T.H. et al., (2017). Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges. *Transportation Research Part C: Emerging Technologies*, 75, pp.197–211.

Rasouli, S., (2016). *Uncertainty in modeling activity-travel demand in complex urban systems*. Eindhoven University of Technology.

Rasouli, S. & Timmermans, H., (2012). Uncertainty in travel demand forecasting models: literature review and research agenda. *Transportation Letters*, 4(1), pp.55–73. Available at: http://www.tandfonline.com/doi/full/10.3328/TL.2012.04.01.55-73.

Recker, W.W., McNally, M.G. & Root, G.S., (1986a). A model of complex travel behavior: Part II—An operational model. *Transportation Research Part A: General*, 20(4), pp.319–330. Available at: http://linkinghub.elsevier.com/retrieve/pii/0191260786900907.

Recker, W.W., McNally, M.G. & Root, G.S., (1986b). A model of complex travel behavior: Part II—An operational model. *Transportation Research Part A: General*, 20(4), pp.319–330. Available at: http://linkinghub.elsevier.com/retrieve/pii/0191260786900907.

Shabanpour, R. et al., (2016). Investigating the applicability of ADAPTS activity-based model in air quality analysis. *Travel Behaviour and Society*.

Shiftan, Y., Kheifits, L. & Sorani, M., (2015). Travel and Emissions Analysis of Sustainable Transportation Policies with Activity-Based Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 2531, pp.93–102. Available at: http://trrjournalonline.trb.org/doi/10.3141/2531-11.

Siyu, L.I., (2015). *Activity-Based Travel Demand Model□: Application And Innovation*. Department of Civil and Environmental Engineering National University of Singapore.

Smith, L., Beckman, R. & Baggerly, K., (1995). TRANSIMS: Transportation analysis and simulation system. *Other Information: PBD: [1995]*, (26295). Available at: http://www.osti.gov/scitech/servlets/purl/88648-fgWOUT/webviewable/.

Statbel, (2017). Land use in Belgium since 1990. Available at: https://statbel.fgov.be/nl/themas/leefmilieu/grond/bodemgebruik#figures.

Usman, M. et al., (2017). Optimal recharging framework and simulation for electric vehicle fleet. *Future Generation Computer Systems*. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167739X17307689 [Accessed July 27, 2017].

Veldhuisen, J., Timmermans, H. & Kapoen, L., (2000). Microsimulation Model of Activity-Travel Patterns and Traffic Flows: Specification, Validation Tests, and Monte Carlo Error. *Transportation Research Record: Journal of the Transportation Research Board*, 1706, pp.126–135. Available at: http://trrjournalonline.trb.org/doi/10.3141/1706-15.

Weiss, C. et al., (2017). Assessing the effects of a growing electric vehicle fleet using a microscopic travel demand model. *European Journal of Transport and Infrastructure Research*, 17(3).

Zhuge, C. et al., (2017). Sensitivity analysis of integrated activity-based model: using MATSim as an example. *Transportation Letters*, pp.1–11. Available at: https://www.tandfonline.com/doi/full/10.1080/19427867.2017.1286772.

## Authors Bibliographic Details

### 1. Dr. Syed Fazal Abbas Baqueri

Syed Fazal Abbas Baqueri has received his PhD degree in Transportation Sciences, in October, 2018 from Hasselt University, Belgium. After his PhD he has started working as Assistant Professor in Civil Engineering Department of DHA Suffa University, Karachi, Pakistan. His research interests include external trip analysis, microsimulation and activity-based modelling, discrete choice modelling and application of these techniques to solve variety of demand oriented transport research problems. He has (co-) authored 5 publications in reputed international peer-reviewed journals.

### 2. Dr. Muhammad Adnan

Dr. Muhammad Adnan has started working as senior researcher at IMOB, Hasselt University in year 2016. He is managing two major work packages of an EU funded project iSCAPE (SMART control of air pollution in Europe) that involves a consortium of 14 different European universities and institutes. The project also involved establishment of a living lab in 6 European cities including Hasselt, where Dr. Adnan is a lead. Prior to joining IMOB, he worked under Prof. Moshe Ben-Akiva (a distinguished MIT Professor) as postdoctoral research associate in Singapore-MIT Alliance for Research and Technology, where he was heavily involved in development process of state-of-the-art integrated activity-based model. He concluded his PhD study from Institute of transport studies, University of Leeds, UK in 2010. He also (co-)authored over 28 publications in international peer reviewed journals and conferences. At IMOB, he is providing supervision to around 5 PhD students. His main research fields includes integrated modelling within activity-based paradigm, assessment of policy impacts, discrete choice modelling, statistical methods, informational intervention and awareness campaigns design and evaluation.

### 3. Prof. dr. Bruno Kochan

Prof. dr. B. Kochan is assistant professor at Hasselt University and is responsible for the development of activity-based transport models at the Transportation Research Institute (IMOB). He has 10+ years of experience in activity-based modelling research, development and practice. During this period, he worked with many researchers from different countries in order to set up activity-based models, which were all based on his doctoral research. He also (co-)authored over 20 publications in international peer-reviewed journals and took care of the daily supervision of in total 8 doctoral candidates. His main research focus lies in the domains of travel behaviour, mobility, activity-based modelling, modelling decision processes of individuals, traffic demand management policies, policy evaluation and impact assessment.

### 4. Prof. dr. ir. Tom Bellemans

Prof. dr ir. T. Bellemans is professor at Hasselt University and is responsible for the research on travel behavior and transportation data processing at the Transportation Research Institute (IMOB). He has over a decade of experience in activity-based modeling research and practice. During this period, he managed the research and development of the FEATHERS activity-based model. He is one of the co-founders of AbeonaConsult, a consulting company focusing on smart products and services in transportation and traffic safety. Prof. dr. ir. T. Bellemans (co-)authored over 50 publications in international peer-reviewed journals and served as the (co-)promotor of over a dozen research projects, several of which were funded by public sector bodies. His main research interests include travel behavior, mobility, activity-based modeling, traffic demand management policies, policy evaluation and impact assessment, and research on and development of tools to support mobility of persons with disabilities.

**Authors Photgraphs**

| | |
|---|---|
| **Dr. Fazal Abbas Baqueri** |  |
| **Dr. Muhamamd Adnan** |  |
| **Prof.dr. Bruno Kochan** | |
| **Prof. Tom Bellemans** |  |

**Manuscript Title:**

**Activity-based Microsimulation Model for Medium-Sized Cities Considering External Activity-Travel: Enhancing FEATHERS Framework**

Highlights:

1. Resident's external activity-travel is integrated in the activity-based model using FEATHERS.
2. Destination choice models are enhanced for locations outside study area by selecting a catchment area.
3. The framework helps application of activity-based model for medium-sized cities.
4. Developed model is applied and validated for three medium-sized cities in Belgium.