sustainability

MDPI

Article

A Methodological Approach for Enriching Activity–Travel Schedules with In-Home Activities

Feng Liu *, Tom Bellemans, Davy Janssens, Geert Wets and Muhammad Adnan *

Transportation Research Institute (IMOB), Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium; tom.bellemans@uhasselt.be (T.B.); davy.janssens@uhasselt.be (D.J.); geert.wets@uhasselt.be (G.W.)
* Correspondence: feng.liu@uhasselt.be (F.L.); muhammad.adnan@uhasselt.be (M.A.)

Abstract: In-home activities are inevitably important parts of individuals' daily schedules, as people spend more time working and doing various other activities (e.g., online shopping or banking) at home. However, conventional activity-based travel demand models (ABMs) only consider travel and travel-related out-of-home activities, ignoring the interaction between in-home and out-of-home activities. To fill in this gap and increase the understanding of what people do at home and how in-home and out-of-home activities affect each other, a new method is proposed in this study. The approach predicts the types and durations of in-home activities of daily schedules generated by ABMs. In model building, statistical methods such as multinomial logit, log-linear regression, and activity sequential information are utilized, while in calibration, the Simultaneous Perturbation Stochastic Approximation (SPSA) method is employed. The proposed method was tested using training data and by applying the approach to the schedules of 6.3 million people in the Flemish region of Belgium generated by a representative ABM. Based on the statistical methods, the mean absolute errors were 0.36 and 0.21 for predicting the number and sum of the durations of in-home activities (over all types) per schedule, respectively. The prediction obtained a 10% and 8% improvement using sequential information. After calibration, an additional 60% and 68% were gained regarding activity participation rates and time spent per day. The experimental results demonstrate the potential and practical ability of the proposed method for the incorporation of in-home activities in activity–travel schedules, contributing towards the extension of ABMs to a wide range of applications that are associated with individuals' in-home activities (e.g., the appropriate evaluation of energy consumption and carbon emission estimation as well as sustainable policy designs for telecommuting).

Keywords: in-home activities; multinomial logit; log-linear regression; activity sequential information; calibration; simultaneous perturbation stochastic approximation (SPSA); activity–travel schedules

Citation: Liu, F.; Bellemans, T.; Janssens, D.; Wets, G.; Adnan, M. A Methodological Approach for Enriching Activity–Travel Schedules with In-Home Activities. *Sustainability* **2024**, *16*, 10086. https://doi.org/10.3390/su162210086

Academic Editor: Gianpiero Greco

Received: 24 October 2024
Revised: 16 November 2024
Accepted: 17 November 2024
Published: 19 November 2024

1. Introduction

Climate change phenomena represent one of the most severe threats to human well-being and sustainable development [1,2]. To address the challenge of climate change, the I-CHANGE (Individual Change of Habits Needed for Green European transition) H2020 EU project [3] has been initiated, aimed at engaging citizens and promoting sustainable activity–travel behavior and lifestyles (i.e., choices that generate less energy consumption and carbon emissions). In this context, activity-based travel demand models (ABMs) come into play. ABMs are a type of highly disaggregated (agent-based) modeling framework that captures individuals' daily activity–travel routines [4,5]. Based on ABMs, the effects of activity–travel behaviors (i.e., out-of-home activities and associated travel) on energy consumption and carbon emissions can be estimated, and a simulation of policies/scenarios that could motivate the shifts to sustainable lifestyle and activity patterns can be produced.

Nevertheless, most of the operational ABMs only consider travel and travel-related out-of-home activities (i.e., activities conducted outside the home); in-home activities (i.e., activities performed at home) are ignored in the process of scheduling individuals' activities [6,7]. Consequently, in-home activities are not contained in the daily schedules generated by ABMs (ABM schedules), leading to incomplete activity sequences. However, according to the US Bureau of Labor Statistics [8], on average, people in the United States spend 1076 min (75% of the day) at home and only 69 and 295 (min) (5% and 20%) on travel and out-of-home activities each day. Thus, while ABMs provide a more advanced framework for modeling travel and travel-related activities, they cannot evaluate individuals' activity behaviors throughout the day. To provide better information and policy recommendations for behavior changes, a complete picture of time use patterns (activity sequences) should be considered that accommodates both in-home and out-of-home activities.

In addition to evaluating behavioral changes, having a clear vision of time use patterns can be helpful in other areas. The first utilization is related to the COVID-19 pandemic. During this period, the outbreak drastically reduced economic activities and population mobility as a result of lockdown measures, leading to a significant drop in visits to workplaces, shops, and recreation locations, with a substantial increase in activities at home [9]. These responses have been anticipated as a window of opportunity for long-term behavioral changes in working and mobility patterns [10]. However, to attain the best perception of the potential results of policies that enhance these sustainable behavioral changes, it is essential to understand how individuals plan and schedule in-home activities to spend the required time on obligatory activities (e.g., work and education), as well as how in-home activities potentially replace their equivalent out-of-home ones (e.g., the reduction in visiting shops as a result of e-shopping activities). The second utilization lies in the area of energy supply chain designs. The increasing trend of doing activities at home has impacted how residential buildings are used and how long people spend on various activities in their homes. This leads to considerable changes in the occupancy patterns of residential buildings compared to pre-pandemic norms. The availability of complete time use patterns can provide necessary inputs for designing more efficient and sustainable supply chains (grids) for water, gas, and electricity by associating people's in-home activities with the amount of these resources they consume [11]. From a sustainability perspective, optimized supply chains contribute to energy conservation and emission reduction by reducing overproduction and resource losses. Demand-focused grid design integrates renewable energy sources, aligning variable supply with usage patterns. This supports innovative building development with energy-efficient systems.

In sum, the traditional ABMS are inadequate for meeting the requirements for the applications mentioned above and addressing the challenges that the post-COVID-19 pandemic world is facing, and further development of the models is required. This necessitates additional research to develop an effective method to enrich ABM schedules with in-home activities to provide a complete picture of time use patterns in and outside homes.

2. Literature Review

2.1. Activity-Based Travel Demand Models

Activity-based models originate from the time–space prism theory presented by Hägerstrand [12], where individuals perform their daily activities bound by time and space constraints [13]. These models (ABMs) predict travel patterns in terms of where, when, how, why, and with whom individuals travel and estimate the travel demand (e.g., the number of vehicles moving between different locations) [4,14]. The central premise of ABMs is the treatment of travel as a derived demand for activity participation. In this modeling framework, travel surveys that record the daily activity–travel sequences of a small sample of respondents during one or a few days, along with land use, transportation network, and socio-economic data, are used to estimate models based on statistical and machine learning techniques [4,6]. For example, [4] used a decision tree modeling approach, and [12] employed statistical models within their ABM framework. Figure 1 depicts different components of a typical integrated ABM framework. The core modeling components are dark grey blocks that require specific inputs, which are given as white blocks. The primary outputs from the model are provided as light grey blocks. A typical ABM model consists of an activity generation or daily pattern generation component and an activity scheduling component. The activity generation component mainly contains the models that represent the choice of a number of activities, their sequence, and several tours (home-based tours) the individuals will perform in a day. The activity scheduling component consists of multiple choices involving activity and travel decision-making, such as time choice, mode choice, and location choice. Usually, a sequential approach is used to model such decisions [15]. This results in individuals' full-day schedules of all trip information. The individual trip information is then processed into an origin–destination matrix, and then fed to a supply-side model (dark grey block) that provides output through traffic flows and travel times on the road network. The outputs of the ABM model, which are a list of activities and their schedule (sometimes referred to as plans), are fed to the traffic assignment model. The traffic assignment model generates a new set of travel times (skim matrices), which serve as input to the ABM for the next iteration.

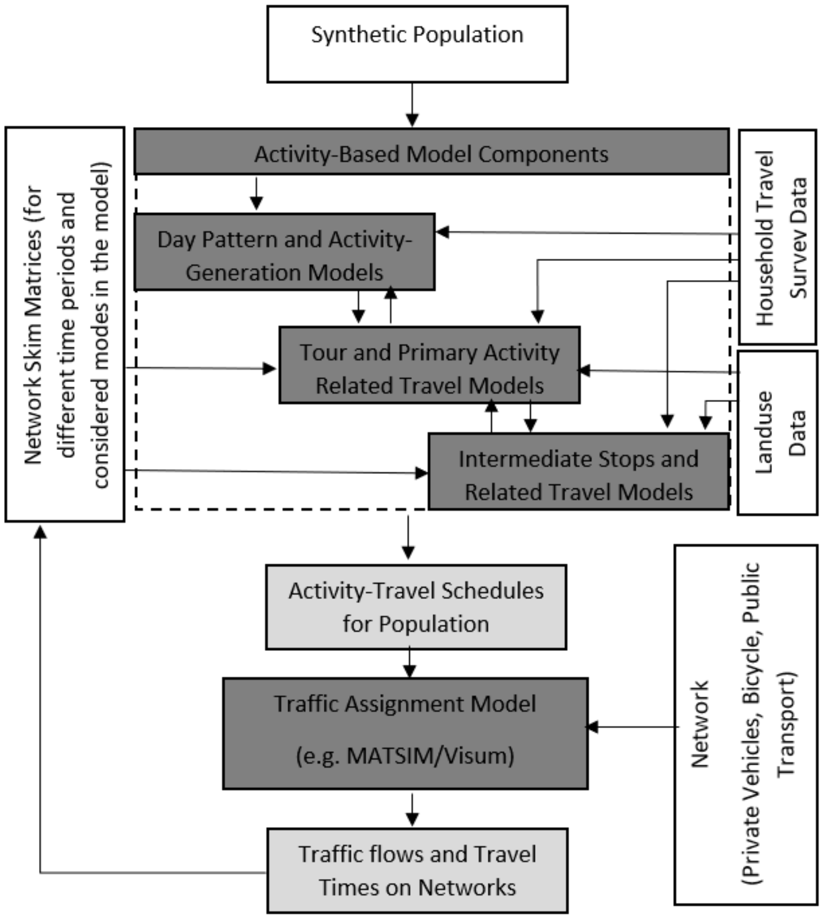

Figure 1. Integrated activity-based model framework with required Modeling and simulation inputs. (Core model components → dark grey blocks, inputs → white blocks, outputs → light grey blocks).

However, while ABMs concentrate on modeling travel and travel-related elements (e.g., travel modes and activities performed at travel destinations), less attention has been given to analyzing and simulating in-home activities. The role of in-home activities in the process of planning and scheduling of individuals' daily activities has been disregarded. This can be attributed to two major reasons: (1) the fact that in-home activities are not directly involved with trips, and (2) the scarcity of data sources that provide required details on planning and scheduling in-home activities. Consequently, in-home activities are not accommodated in ABM schedules, leading to incomplete daily activity sequences. For instance, a typical ABM schedule would be 'activities at home at 4–8 h→ working in the office at 9–17 h → shopping in supermarkets at 17:20–18:20 h → activities at home at 19–4 h (next day)', with travel being implicit between every two consecutive locations. According to this schedule, the person conducted three trips and two out-of-home activities (i.e., work and shopping), while staying at home for two periods (i.e., 4–8 h and 19–4 h). However, the specific activities that the person performed during these two periods are unknown. There exist empirical studies that investigated individual time allocation behavior for in-home activities, along with their impact on out-of-home activities, but the incorporation of in-home activities in an operational ABM is scarce in the literature. This research aims to present a methodological framework to further enrich these home periods with detailed in-home activities.

2.2. Factors Affecting In-Home Activities

In-home activities have been studied based on various statistical models, and both individual–household and activity–schedule attributes have been identified as essential factors affecting in-home activity choices and time allocation. Regarding individual-household attributes, Bhat and Koppelman [16] and Yamamoto and Kitamura [17] analyzed individuals' time allocation patterns for weekly in-home and out-of-home discretionary activities, and showed that both individual (e.g., age, gender, and income) and household (e.g., household size and the number of children) variables are important determinants of location choices (between in-home and out-of-home places) and time allocation for discretionary activities. Meanwhile, Bhat and Gossen [18] identified similar influences of these variables on leisure activity choices in and outside the home. Concerning activity–schedule attributes, Doherty [19] and Shabanpour et al. [7] found that the start time and duration of an activity have a significant impact on the location choice of the activity. Along with that, the attributes of other activities or travel on the same day also influence a particular activity. For instance, Miller and Roorda [20] revealed that the time spent on work or household maintenance affects the location choice of a discretionary activity, while Yamamoto and Kitamura [17] recognized that individuals who spend more time commuting tend to perform more activities at home.

The literature also investigates changes in people's activities (both in and outside the home) and travel behavior because of the COVID-19 pandemic. As described in Section 1, during the pandemic, the combination of social distancing and travel restrictions, as well as people's decisions to avoid the infection risk, resulted in an unprecedented change in the mobility styles of individuals [21,22]. Activities, including work, education, and shopping, witnessed a significant shift to online settings, and the total number of in-home activities soared. Among the activities, the shift to working from home (WFH) was a significant change in people's routines [23], and is a long-lasting trend rather than a temporary phenomenon that just occurred during the pandemic [24–26]. The shift towards the new work modality will further result in reduced commuting trips and changes in activity routines [27,28], leading to the emergence of new activity and travel patterns featuring an increasing level of probabilities for in-home activity choices and time allocation. This signals the growing importance of in-home activities in the scheduling of individuals' daily activities and prompts further research.

2.3. In-Home Activity Prediction

While the rich literature described above analyses the characteristics and influencing factors of in-home activities, only a few efforts have focused on the prediction of these activities within an ABM framework. The studies [6,7] upgraded the existing ABM, namely ADAPTS, by integrating in-home activity planning and scheduling into the modeling process, making the new framework capable of simulating in-home activities alongside out-of-home activities. A joint discrete–continuous model consisting of multinomial logit and log-linear regression [7] and a sequential conditional probability approach composed of multinomial logit and hazard-based duration modeling [6] were employed to predict in-home activity types and duration. Both of these studies employed time use survey data alongside household travel data to estimate behavioral models. Six general in-home activity types were used and the prediction focused on activity episodes (i.e., the type and duration of each activity instance being estimated), while the derived results were a sequence of in-home activities in the temporal order. Khalil and Fatmi [29] utilized machine learning techniques; six learning methods were considered, including artificial neural networks, regression trees, ensembles, support vector machines, k-nearest neighbor, and Gaussian process regression. They considered four activity types; the prediction concentrated on the occurrence of each activity type in the schedule, while the obtained results were a two-dimensional vector for each type, indicating whether the corresponding activity type occurs (with at least one instance) and how long the activity would be (if it occurs). Hesam et al. [30] presented their activity-based modeling framework, namely SALT, which only incorporated two types of in-home activities i.e., home chores and home leisure. Their modeling framework employed behaviorally based econometric, machine-learning, and data-mining techniques and used the Halifax Space Time Activity Research (STAR) household survey.

In addition to the methods adopted, differences and shortcomings exist between the studies mentioned in the above paragraph. These are given below:

- In studies [6,7], six general in-home activity types were used; in comparison, study [29] predicts only four types of in-home activities, and study [30] incorporates only two in-home activity types. Given that there may be several activity episodes of the same type performed at different times of the day, the studies [6,7] provide more detailed information on activity patterns than the third study, such as the total number of episodes of each type and the start times of these episodes, as well as the sequential order of episodes of different types.
- None of these studies conducted additional analysis to examine how the predicted results perform in the context of daily activity sequences that are under a specific sequential constraint. As standard statistical models or machine learning methods offer an effective technique for modeling each single activity episode, they discard the details of activity ordering and transitions embedded in activity and travel patterns. When the predicted activities are filled into an individual's daily schedule, the activity patterns (composed of both in-home and out-of-home activities) should follow a specific sequential constraint, i.e., the choice of activities is dependent on the preceding activity engagement [31]. For instance, if a sports activity is carried out (either in or outside the home) in the morning, there is a slight chance that it is performed again in the evening. These interdependencies of daily activities are crucial in activity–travel decision-making.
- While time use surveys, which are commonly used for in-home activity modeling [6,7,29,30], provide valuable information on activity–travel behavior, they have an intrinsic weakness (e.g., the lengthy period of data processing and under-reporting of short-duration or infrequent activities) [31]. This leads to a certain level of deviation between the predicted results and the ones that reflect the actual behavior of the current situation. Similar problems have arisen from the OD matrix generated by ABMs, in which differences have been identified between the derived matrix (along with its assignment to road networks) and the actual traffic flow patterns. To reduce these discrepancies, a calibration process has been used to match the predicted OD as closely as possible to the actual measurements collected from the road network. However, in the previously described in-home activity prediction methods, no similar (calibration) processes were considered to update the prediction results to make the marginal distributions of the estimated activities well aligned with the observed ones and to generate the prediction results as an accurate representation of current in-home activity behavior.

2.4. Research Contributions

Extending the current research on predicting in-home activities while addressing the limitations mentioned above, this study proposes a new approach that integrates statistical models, activity sequential information, and calibration into the prediction framework. The goal of the approach is to predict the types and durations of activities for each given home period of an ABM's predicted activity–travel schedule. Compared with the existing prediction methods, the new approach offers significant advantages. (1) It is based on integrating conventional statistical models and activity sequential information inherent to human activity and travel behavior. (2) A calibration process is employed to better match the predicted results to the observed marginal activity distributions (e.g., the percentage of people who perform at least one activity of a given type each day and the average time spent on the activities per person). (3) By calibration, the proposed method can be easily transferred to other application regions with marginal activity data and similar conditions to the originally model-developed area (regarding road networks and social-economic development) which do not have detailed time use surveys. (4) The method is tested by applying the approach to the schedules (around 6.3 million) of individuals in the Flemish region of Belgium generated by a representative ABM, and the potential and practical ability of the approach is demonstrated.

3. Methodological Approach for the Enrichment Method

3.1. Overall Structure of the Enrichment Method

The method proposed here consists of two major parts: model building and model calibration, as outlined in Figure 2. The Model Building Section (namely the MB process, Section 3.3) is further divided into four steps, including (1) data processing, (2) the prediction of activity types and durations of each single in-home activity record, (3) the prediction of each given home period (HomeP), and (4) prediction enhancement based on activity sequential information (ASI). In the Model Calibration Section (denoted as the MA calibration process, Section 3.4), the previously derived models are further improved based on observed marginal activity (MA) data.

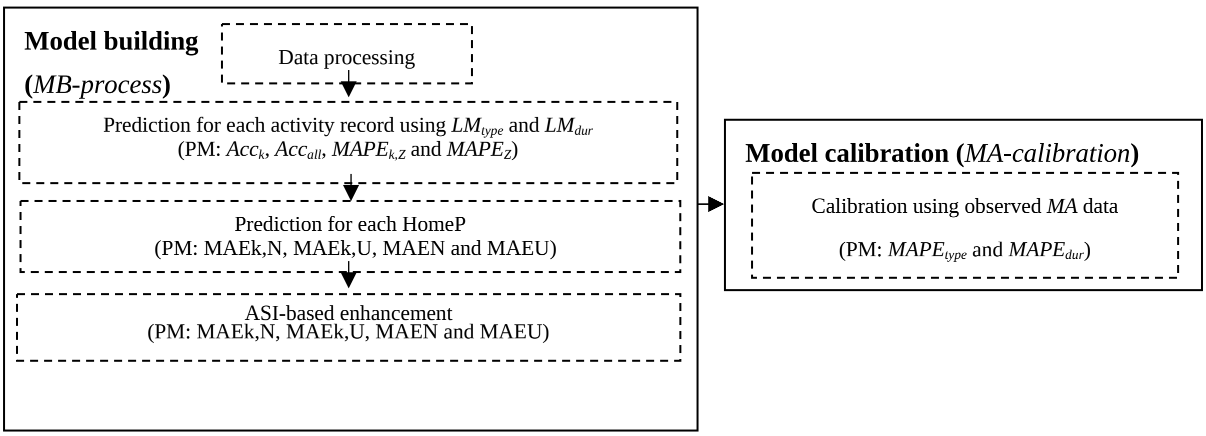

Figure 2. The overall structure of the enrichment method. Note: PM denotes the prediction performance measures.

3.2. The Training Data

The data for training and testing the model were obtained from the 2019 American Time Use Survey (ATUS) (U.S. Bureau of Labour Statistics, 2019) [8], which was administered to obtain detailed information about the time use patterns of a sample of 9435 individuals along with their socio-economic conditions. The data were collected for a full schedule of a random day, including every in-home, out-of-home, and travel activity performed by the respondent during that day. The U.S. data from 2019 reflect pre-COVID-19 conditions. This means they capture typical household behaviors, occupancy patterns, and daily activity distributions as they existed before the significant lifestyle changes prompted by the pandemic. Table 1 illustrates the activity records of a respondent on a Thursday; all the activities are arranged according to the temporal order, forming an observed schedule (i.e., ScheO). The detailed categories of the in-home activities of the survey provide valuable information for studying these activities. The original data featuring weekdays from 4642 individuals were extracted into a dataset (i.e., ATUS-wd), which was further divided into training and test sets with 80% and 20% of the individuals for model estimation and validation, respectively. For each day, the continuous time interval for doing (one or a few) activities at home is defined as a home period (i.e., HomeP). For instance, Table 1 shows two HomePs spanning 4:00–7:15 and 20:50–4:00 (next day), respectively.

Table 1. Activity records of a respondent on a single day.

Start Time	End Time	Activity Duration	Activity Type *	Location
4:00	6:30	150	S	Home
6:30	7:15	45	P	Home
7:15	8:20	65	Travel	/
8:20	13:00	280	Mo	Office
13:00	14:00	60	Po	Office
14:00	19:30	330	Mo	Office
19:30	20:15	45	Travel	/
20:15	20:50	35	H	Home
20:50	21:50	60	P	Home
21:50	23:00	70	L	Home
23:00	4:00	300	S	Home

* Activity types can be referred to in Table 2.

Table 2. In-home activity types *.

Activity Type	Le	Definition	Sub-Category	Le_o
Sleep	S	Sleep and rest at night and during the daytime	1 (100%)	
Personal care	P	Washing, dressing, and grooming (1); eating and drinking (11)	1 (52%), 11 (48%)	P_o
Household	H	Household maintenance (2), and caring for household (3) and non-household members (4)	2 (79%), 3 (20%), 4 (1%)	H_o
Leisure	L	Socializing, relaxing, and leisure (12); sports and exercise (13)	12 (98), 13 (2%)	L_o
Discretionary	D	Online shopping (7), personal care (8), and household (9) services; religious (14) and volunteer (15) activities	7 (18%), 8 (4%), 9(6%), 14 (50%), 15 (22%)	D_o, SH_o
Mandatory	M	Work and work-related activities (5) and education (6)	5 (79%), 6 (21%)	M_o

*: The columns from the left to right denote the in-home activity types, represented letters, definitions, sub-categories, and corresponding letters for out-of-home activities. In 'Definition', the number in the bracket indicates the sub-categories, while in 'Sub-category', the value in the bracket represents the share of the sub-category among the total activity records of the corresponding type.

3.3. Model Building (MB Process)

3.3.1. Data Processing

Six commonly used activity types, including sleep (S), personal care (P), household (H), leisure (L), discretionary (D), and mandatory (M) activities were considered [7]. The definition of these types, along with the sub-categories of each type [32], is described in Table 2. To explore the sequential correlations of in-home activities with out-of-home ones, activities performed outside the home were also utilized. These activities were divided into personal care (P_o), household (H_o), leisure (L_o), mandatory (M_o), discretionary (D_o), and shopping (SH_o) activities, with the former four types being equivalent to P, H, L, and M, while the latter two (D_o and SH_o) originate from D.

The explanatory variables (presented in Table 3) include information on the classification of the attributes (such as individual, household, and activity schedules) to account for the impact of these variables on the prediction. The survey adopts the classification of individual and household variables, while the decision tree method [33] was used for activity–schedule variables. This approach was employed to choose the most significant cutting points for each variable, such that the response values (i.e., in-home activity types) were as similar as possible within each obtained interval while as different as possible across intervals. In addition to the explanatory variables, the other major variables used in the process are listed in Table A1 in the Appendix A.

Table 3. Explanatory variables.

Variable Name	Variable Classification
Individual variables (4)	
Work	Employment status; 1: full-time, 2: part-time, and −1: unemployed
Edu	Student status; 1: full-time student, 2: part- time student, and −1: not a student
Age	Age; 1: 0–12, 2: 13–17, 3: 18–34, 4: 35–54, 5: 55–64, 6: 64–74, and 7: 75+
Sex	Gender; 1: male, and 2: female
Household variables (6)	
WoS	Employment status of spouse; 1: full-time, 2: part-time, and −1: unemployed, and 0: no spouse
Nchi	Number of children under the age of 18; 0: none, 1: one, 2: two, and 3: 3+
Nper	Number of people in the household; 1: one (the respondent alone), 2: two, and 3: 3+
Earn	Weekly earnings; 0: none, 1: USD 0–1000, 2: USD 1000–2000, and 3: USD 2000+
Achi	Age of the youngest child under the age of 18; 0: none, 1: 0–12, and 2: 13–17
Aold	Age of the oldest person; 1: 18–34, 2: 35–54, 3: 55–64, 4: 65–74, and 5: 75+
Activity–schedule variables (26)	
ActST	Start time of the predicted time interval; 1: 0–4:00 a.m., 2: 4:00–6:15 a.m.; 3: 6:15–9:10 a.m., 4: 9:10 a.m.–16:30 p.m., 5: 16:30–19:25 p.m., 6: 19:25–20:50 p.m., 7: 20:50–21:50 p.m., and 8: 21:50–24:00 p.m.
PerD	Duration (min) of the predicted home period, for which one or several activities are to be predicted; 1: 0–75, 2: 75–520, 3: 520–670, and 4: 670+
Out1, Out2, Out3, Out4, Out5, Out6	Total duration (min) of out-of-home activities of each type; Out1 (P_o): 1: 0–15, 2: 15–105, 3: 105+; Out2 (H_o): 1, 0–15, 2: 15–215, 3: 215+; Out3 (L_o): 1: 0–15; 2: 15–215; 3: 215+; Out4 (D_o): 1: 0–15; 2: 15–170; 3: 170+; Out5 (M_o): 1: 0–15; 2: 15–80; 3: 80–480; 4: 480+; and Out6 (SH_o): 1: 0–15; 2: 15–200; 3: 200+
Trip1, Trip2, Trip3, Trip4	Total duration for each mode of travel; Trip1 (car driver): 1: 0–15; 2: 15–150; 3: 150+; Trip2 (car passenger): 1: 0–65; 2: 65–240; 3: 240+; Trip3 (walking or biking): 1: 0–60; 2: 60+; and Trip4 (bus, subway, or train): 1: 0–15; 2: 15+
OutT, TripT	Total duration for all out-of-home activities and trips; OutT: 1: 0–315; 2: 315–695; 3: 695+; TripT: 1: 0–45; 2: 45–180; 3: 180+

Table 3. *Cont.*

Variable Name	Variable Classification
In1, In2, In3, In4, In5, In6	Total duration for each type of in-home activities conducted before the predicted time interval; In1 (S): 1: 0–10; 2: 10–275; 3: 275+; In2 (P): 1: 0–10; 2: 10–65; 3: 65+; In3 (H): 1: 0–10; 2: 10–80; 3: 80+; In4 (L): 1: 0–150; 2: 150–400; 3: 400+; In5 (D): 1: 0–10; 2: 10–95; 3: 95+; and In6 (M): 1: 0–10; 2: 10–120; 3: 120+
Pre1, Pre2, Pre3, Pre4, Pre5, Pre6	Duration of the preceding in-home activity before the predicted time interval; Pre1 (S): 1: 0–60; 2: 60–85; 3: 85+; Pre2 (P): 1: 0–20; 2: 20–160; 3: 160+; Pre3 (H): 1: 0–10; 2: 10–190; 3: 190+; Pre4 (L): 1: 0–35; 1: 35–120; 3: 120–245; 4: 245+; Pre5 (D): 1: 0–30; 2: 30+; and Pre6 (M): 1: 0–135; 1: 135+

3.3.2. Prediction for Each Activity Record

Various statistical methods have been explored to model in-home activity choices and time allocation. Among the methods, the multinomial logit model (MNL) is the most basic form of discrete choice models and provides a closed form and efficient computation for choice probabilities. This model is particularly suitable for large-scale activity prediction [6,7]. Given the potential application of the proposed approach to a large area, as well as the multiple runnings of an iterative process in the calibration, the computation time is regarded as an important factor. Thus, the MNL was adopted in the approach. In terms of activity durations, the log-linear regression model [7] was chosen to estimate the durations of activities of each type.

Specifically, K is the total number of in-home activity types, M is the total number of explanatory variables, and x_m and C_m (m = 1, . . . , M) are each of the explanatory variables and the number of categories of this variable, respectively. Moreover, Y is the dependent variable for in-home activity types, and $Pr(Y = k)$ (k = 1, . . . , $K - 1$) and $Pr(Y = K)$ are the probabilities of $Y = K$ and $Y = K$. The log ratio (i.e., logit) between $Pr(Y = k)$ and $Pr(Y = K)$ for an individual is modeled as follows.

$$\begin{aligned} &\log \frac{\Pr(Y=k)}{\Pr(Y=K)} = \beta_{k,0} + \beta_{k,1}\cdot x_1 + \ldots + \beta_{k,M}\cdot x_M + \varepsilon_k \; (k = 1, \ldots, K-1) \\ &\sum_{k=1}^{K} \Pr(Y=k) = 1 \end{aligned} \tag{1}$$

where, $\beta_{k,0}$ and $\beta_{k,1}$, . . ., and $\beta_{k,M}$ are the intercept and slope parameters; $^{\epsilon_k}$ is the random error term corresponding to unobserved factors, with a standard type-I extreme value distribution.

Similarly, for a given activity type k, Z is the dependent variable for activity duration; the log of Z is characterized as follows:

$$\log(Z|k) = \alpha_{k,0} + \alpha_{k,1}\cdot x_1 + \ldots + \alpha_{k,M}\cdot x_M + \Phi_k \tag{2}$$

where, $\alpha_{k,0}$ and $\alpha_{k,1}$, . . ., and $\alpha_{k,M}$ are the intercept and slope parameters, respectively; while Φ_k is the random error depicting unobserved factors and is assumed to have a normal distribution. Note that in Equations (1) and (2), a linear function of the explanatory variables is considered, and the error terms $^{\epsilon_k}$ and Φ_k are identically distributed across individuals [12,13]. Moreover, there are $(M + 1) \times (K - 1)$ and $(M + 1) \times K$ parameters in the two equations, respectively. The models containing all the parameters which are estimated using the in-home activity records in the training set are denoted as LM_{type} and LM_{dur}. LM_{type} contains only one model characterizing the probabilities of all the activity types, whereas LM_{dur} is composed of six independent sub-models for the prediction of the durations of different activity types.

3.3.3. Prediction for Each Given Home Period (HomeP)

Based on LM_{type} and LM_{dur}, the activities of each given home period *HomeP* are predicted. To this end, a home period prediction method, namely the *HomePPM* method (see Figure 3), is designed to estimate the activity type and duration of each possible in-home activity in *HomeP*. Specifically, t_s, $t_{e,}$ and Int_{home} are the start time, end time, and the time interval of *HomeP*, with $Int_{home} = t_e - t_s$; r, $\hat{Y}_r$, and $\hat{Z}_r$ are the order, activity type, and duration of the predicted activity in *HomeP*; and t and Int_{rem} are the start time and remaining interval of *HomeP*, respectively. Moreover, Dur_{min} is defined as the threshold for the minimum duration of activities. This process begins with $r = 1$, $t = t_s$ and $Int_{rem} = Int_{home}$, and $\hat{Y}_r$ and $\hat{Z}_r$ are obtained based on LM_{type} and LM_{dur}, respectively. Afterwards, Int_{rem} is updated with $Int_{rem} \leq Int_{rem} - \hat{Z}_r$ and compared against Dur_{min}. If $Int_{rem} \geq Dur_{min}$, this process is repeated for the prediction of the next activity (i.e., $r \leq r + 1$ and $t \leq t + \hat{Z}_r$) using the updated values of the predictors. Otherwise, if $Int_{rem} < Dur_{min}$, the remaining interval is not sufficiently long for a possible activity; this process stops, and Int_{rem} is assigned to $\hat{Z}_r$ (i.e., $\hat{Z}_r \leq \hat{Z}_r + Int_{rem}$). Note that only some of the activity–schedule variables (e.g., ActST, PerD, In1-In6, and Pre1-Pre6) are renewed in the update of the predictors. The values of the other variables (e.g., the individual and household variables) remain identical (for the same schedule). The final output of this process is a set of estimated in-home activities with the predicted types and durations performed within *HomeP*. A daily schedule may contain several *HomePs*; the predicted in-home activities of all the periods, along with the original out-of-home activities in the schedule, form a complete (predicted) daily sequence according to the temporal order (i.e., *ScheI*).

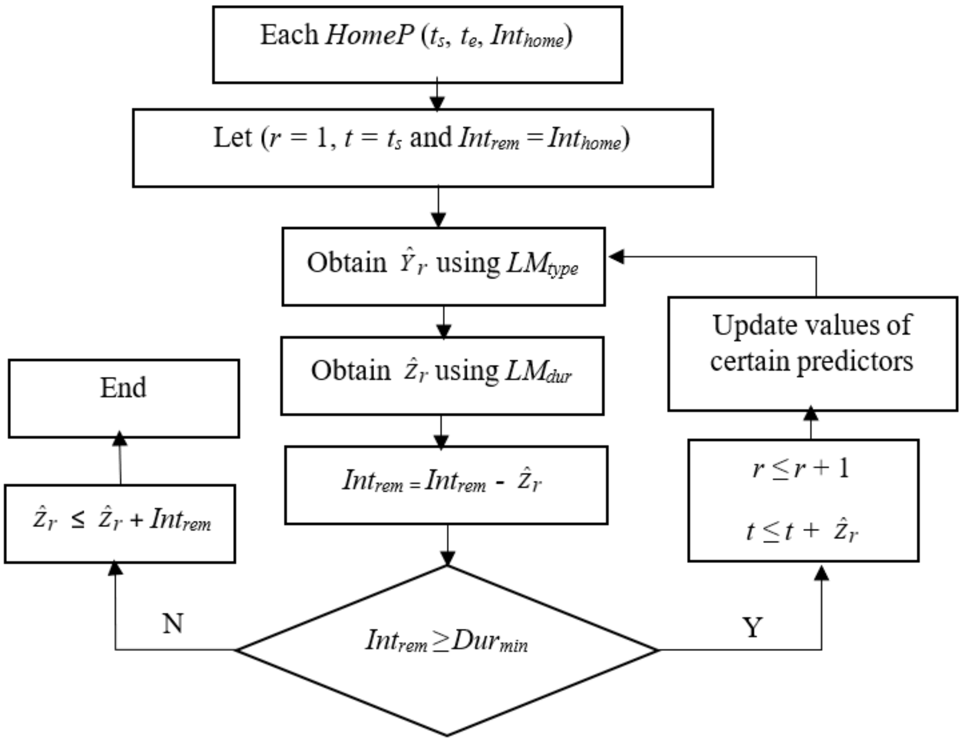

Figure 3. The home period prediction method (*HomePPM*).

Differences exist between the models (LM_{type} and LM_{dur}) and *HomePPM*. LM_{type} and LM_{dur} predict each given activity record, whereas *HomePPM* handles an additional uncertain factor, which is the number of activities possibly contained in *HomeP*. Thus, *HomePPM* provides the estimation of three elements, including the number of activities in *HomeP* as well as the activity type and duration of each of these activities.

3.3.4. Prediction Enhancement Based on Activity Sequential Information (ASI)

A method (i.e., the ASI-based method) was developed to take the previously obtained schedules (ScheIs) as well as the sequential information (derived from the survey data) as inputs and aimed to generate improved inferences. Using GPS and mobile phone data, a similar process was utilized to improve the machine learning results on travel modes and out-of-home activities [34]. This method can be illustrated by the predicted activity–travel schedule of a respondent depicted in Figure 4. According to the observed data, the respondent has conducted the sequence of activities, 'S-H-P-Mo-M-P-L-H-S.' However, based on the results of *HomePPM*, the predicted sequence is 'S-H-P-M_o-*L*-P-L-H-S'; the activity of M at the fifth position in the observed sequence is wrongly forecasted in the predicted schedule as L (i.e., because the predicted probability $Pr(Y = \text{'}L\text{'}) = 0.31$ is the maximum probability across all the activity types). To identify and possibly correct this error, the probability $Pr(\cdot)$ for each in-home activity in the predicted sequence was examined. If $Pr(\cdot)$ is smaller than a threshold TH_1 (e.g., 0.5 in this experiment), the corresponding activity (denoted as A_{fal}) is assumed to have a high likelihood of being a false inference, and an enhancement process is applied to this activity to improve its prediction in the following manner.

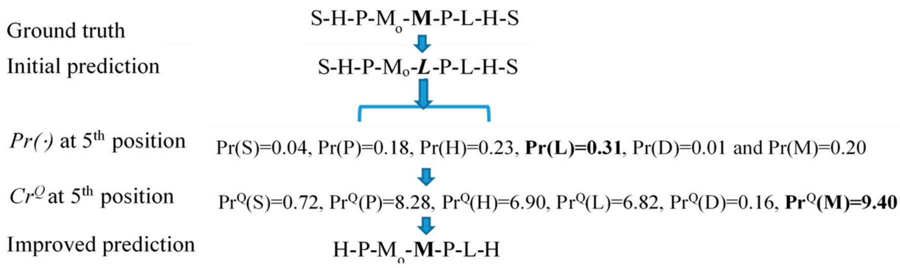

Figure 4. An illustration of the ASI-based method. Note: in *Pr* and Cr^Q, the dependent variable Y is omitted, and the maximum value for *Pr* or Cr^Q is in bold; Cr^Q has units of 1×10^{-7}.

1. If A_{fal} is adjacent to an out-of-home activity (e.g., M_o at the fourth position of the predicted sequence), this out-of-home activity is regarded as a reference activity (referred to as A_{ref}).
2. Otherwise, if A_{fal} (e.g., L at the seventh position) is in the middle of a set of consecutive in-home activities and does not neighbor any out-of-home activities, a second in-home activity (e.g., P or H at the sixth or eighth position) is selected if it is adjacent to A_{fal} and has $Pr(\cdot)$ exceeding a threshold TH_2 (e.g., 0.9). The selected activity is considered to be a possibly correct prediction and is used as A_{ref} to fix the potentially false inference of A_{fal}.
3. The above obtained A_{ref} (from step 1 or 2) is used to compute the correction factor Cr^Q, and the activity (denoted as A_{mod}) with the maximum value of Cr^Q is chosen as the revised activity of A_{fal}.
4. After the revision, the duration of the new activity A_{mod} is re-estimated using LM_{dur} for the corresponding type.
5. In case A_{ref} is not found (i.e., no activities appear in the adjacent area which are either out-of-home activities or in-home activities with a high probability), the revision is not performed.

With the appropriate thresholds TH_1 and TH_2, it is more likely to correct the false prediction while maintaining accurate inference results.

The sequential information is represented in a transition probability matrix between different activity types, e.g., 6×6 in this study. The correction factor Cr^Q is a combination of the sequential information and the predicted probability previously obtained from LM_{type}. Specifically, k_1 and k_2 ($k_1, k_2 = 1, \ldots, 6$) are the types of activities performed consecutively according to the temporal order (i.e., 'k_1-k_2'), and $Tr(k_2 \mid k_1)$ is the transition factor from k_1 to k_2. $Tr(k_2 \mid k_1)$ can be calculated as follows.

$$Tr(k_2|k_1) = \frac{F(k_2|k_1)}{\sum_{k_1=1}^{6} F(k_2|k_1)} \tag{3}$$

where, $F(k_2 \mid k_1)$ is the observed frequency of k_2 followed by k_1. Based on $Tr(k_2 \mid k_1)$, we derive a factor (i.e., $Cr^{\mathrm{T}}(Y = k2)$) of the succeeding activity with the type of k_2 given the previous activity of k_1 as follows:

$$CrT(Y = k2) = \Pr(Y = k_2) \cdot T_r(k_2|k_1). \tag{4}$$

where $Pr(Y = k_2)$ is the prediction probability obtained from LM_{type}. Based on Equation (4), however, the correction factor of activity is biased towards frequently performed activities (e.g., H and P), as transition probabilities $Tr(k_2 \mid k_1)$ to these activities are likely to be higher than to other less-common activities. Consequently, most of the activities under such modification would be redirected to these frequent types. To avoid this problem, $Tr(k_2 \mid k_1)$ is divided by the observed frequency of the succeeding activity type k_2, resulting in the factor $Qr(k_2 \mid k_1)$.

$$Qr(k_2|k_1) = \frac{F(k_2|k_1)}{\sum\limits_{k_1=1}^{6} F(k_2|k_1) \cdot \sum\limits_{k_2=1}^{6} F(k_2|k_1)} \tag{5}$$

The modified correction factor (i.e., $Cr^{\mathrm{Q}}(Y = k_2)$ is as follows:

$$CrQ(Y = k_2) = \Pr(Y = k_2) \cdot Q_r(k_2|k_1). \tag{6}$$

Similarly, the modified correction factor of the previous activity with the type of k1 (i.e., $Cr^{\mathrm{Q}}(Y = k_1)$ provided so that the succeeding activity has the type of k2 can be obtained as follows:

$$CrQ(Y = k_1) = \Pr(Y = k_1) \cdot Q_r(k_1|k_2) \tag{7}$$

In Figure 4, the transitions from M_o to L and from M_o to M are $Qr(L \mid M_o) = 2.2$ and $Qr(M \mid M_o) = 4.7$ (see corresponding case study Section 4.4 b), respectively, leading to $Cr^{\mathrm{Q}}(L)$ (6.82×10^{-7}) being smaller than $Cr^{\mathrm{Q}}(M)$ (9.40×10^{-7}) (calculated by Equation (6)). The initially predicted activity L at the fifth position is thus revised as M after the modification.

3.4. Model Calibration

The Simultaneous Perturbation Stochastic Approximation (SPSA) method is a stochastic gradient approximation algorithm, which is based on an easily implemented and highly efficient gradient approximation that can calibrate a large number of parameters simultaneously using only two measurements of the objective function [35]. The performance of the algorithm for solving large-scale multivariate optimization problems has been documented in several studies [36,37]. Therefore, SPSA was chosen for model calibration.

Specifically, to find the optimal parameters, the algorithm starts from an initial estimation of the parameter vector, and iteratively traces a sequence of parameter estimations which make the objective function converge to a small value based on gradient approximation. The iterative form is as follows.

$$\theta_{n+1} = \theta_n - s_n \cdot \hat{g}_n(\theta_n) \tag{8}$$

$$\hat{g}_n(\theta_n) = \begin{bmatrix} \frac{\mathrm{O}(\theta_n + b_n \cdot \Delta_n) - \mathrm{O}(\theta_n - b_n \cdot \Delta_n)}{2b_n \cdot \Delta_{n1}} \\ \cdots \\ \cdots \\ \frac{\mathrm{O}(\theta_n + b_n \cdot \Delta_n) - \mathrm{O}(\theta_n - b_n \cdot \Delta_n)}{2b_n \cdot \Delta_{nq}} \end{bmatrix} = \frac{\mathrm{O}(\theta_n + b_n \cdot \Delta_n) - \mathrm{O}(\theta_n - b_n \cdot \Delta_n)}{2b_n} \cdot [\Delta_{n1}^{-1}, \Delta_{n2}^{-1}, \ldots, \Delta_{nq}^{-1}]^T \tag{9}$$

$$s_n = \frac{s}{(\mathrm{A} + \mathrm{n} + 1)^{\eta}}, b_n = \frac{b}{(\mathrm{n} + 1)^{\gamma}} \tag{10}$$

In Equation (8), θ_n denotes the estimate of the parameter vector with q-dimensions (i.e., $\theta_n = (\theta_{n1}, \ldots, \theta_{nq})^T$) in the *nth* iteration of the algorithm, $\hat{g}_n(\theta_n)$ (with q-dimensions) is the approximation of the gradient at θ_n, and s_n is a non-negative coefficient controlling the *nth* step size in the updates of θ_n. In Equation (9), $O(\cdot)$ represents the objective function, Δ_n is an q-dimensional perturbation vector (i.e., $\Delta_n = (\Delta_{n1}, \Delta_{n2}, \ldots, \Delta_{nq})^T$), with each component Δ_{ni} ($i = 1, \ldots, q$) being randomly chosen as either 1 or -1 under the same probability of 0.5 (i.e., under a Bernoulli distribution), and b_n is a positive coefficient defining the region where two measurements (i.e., $\theta_n + b_n \cdot \Delta_n$ and $\theta_n - b_n \cdot \Delta_n$) of the objective function are calculated in order to obtain the gradient approximation. In Equation (12), s, A, η, b, and γ are the algorithm parameters, generating s_n and b_n. As n increases, both s_n and b_n become smaller.

The calibration process aims to update the parameter vectors $\beta = (\beta_{k,0}, \beta_{k,1}, \ldots, \beta_{k,M})$ ($k = 1, \ldots, K-1$) and $\alpha = (\alpha_{k,0}, \alpha_{k,1}, \ldots, \alpha_{k,M})$ ($k = 1, \ldots, K$) in LM_{type} and LM_{dur}, respectively, in order to make the deviations between the predicted and observed marginal activity distributions as small as possible. To this end, two object functions (i.e., $O(\cdot)$ in Equation (9)) are defined, including $O_{type}(\beta)$ for activity types and $O_{dur}(\alpha)$ for the average time spent, and they are formulated as follows.

$$\begin{aligned} O_{type}(\beta) &= \sum_{k=1}^{K} |\hat{R}_k - R_k| + \sum_{j=1}^{m} \sum_{c=1}^{C_m} \sum_{k=1}^{K} |\hat{R}_{k,m,c} - R_{k,m,c}| \\ O_{dur}(\alpha) &= \sum_{k=1}^{K} |\hat{T}_k - T_k| + \sum_{j=1}^{m} \sum_{c=1}^{C_m} \sum_{k=1}^{K} |\hat{T}_{k,m,c} - T_{k,m,c}| \end{aligned} \tag{11}$$

where, R_k, $R_{k,m,c}$, T_k, and $T_{k,m,c}$ are the observed marginal activity variables (i.e., MA variables); R_k and $R_{k,m,c}$ are defined as the percentage (participation rate) of the population and of people with a certain category of $x_m = c$ who perform at least one activity of type k each day, while T_k and $T_{k,m,c}$ are the average time spent on the activities per person per day over the population and this group, respectively. $\hat{R}_k$, $\hat{R}_{k,m,c}$, $\hat{T}_k$, and $\hat{T}_{k,m,c}$ are the corresponding predicted MA variables, and obtained from the estimated schedules according to Equation (12).

$$\begin{aligned} \hat{R}_k &= \frac{\sum_{d=1}^{D} Ind(\hat{N}_{k,d} > 0)}{D}, \hat{T}_k = \frac{\sum_{d=1}^{D} \hat{U}_{k,d}}{D} \\ \hat{R}_{k,m,c} &= \frac{\sum_{d=1}^{D_{m,c}} Ind(\hat{N}_{k,d} > 0)}{D_{m,c}}, \hat{T}_{k,m,c} = \frac{\sum_{d=1}^{D_{m,c}} \hat{U}_{k,d}}{D_{m,c}} \end{aligned} \tag{12}$$

In Equation (12), d represents a schedule; D and $D_{m,c}$ are the total numbers of schedules of all individuals and of individuals with $x_m = c$; while $\hat{N}_{k,d}$ and $\hat{U}_{k,d}$ are the predicted number and sum of durations of activities of k in d, respectively. Ind $(\cdot)$ is a Boolean function, being equal to 1 if $\hat{N}_{k,d} > 0$ and 0 if otherwise. In the calibration process, if there are small changes in $O_{type}(\beta)$ and $O_{dur}(\alpha)$ for several successive iterations or if the maximum allowable number of iterations (denoted as TH_{ite}) is reached, the process terminates.

4. Case Study

In this section, the performance of the MB process was examined using the test set, in terms of the prediction accuracy (or errors) in each step (including LM_{type}, LM_{dur}, *HomePPM*, and the ASI-based method) of the process.

4.1. Composition of In-Home Activities

Out of all the activity records, in-home activities account for the largest share of 63% while out-of-home activities and travel only undertake 19% and 18%, respectively. In addition, large variations are observed across all the in-home activities, with household (H), personal care (P), leisure (L), and sleep (S) displaying high frequencies of 19%, 16%, 14%, and 12%, while the remaining discretionary (D) and mandatory (M) activities have merely a small proportion of 1% each. Similarly, large discrepancies exist in activity durations, with S having the longest average duration of 292 min, M and L possessing the median of 112 and 99 (min), while D, H, and P are featured with the shortest durations of 47, 39, and 29 (min), respectively. Table 4 lists the mean, minimum (min), and maximum (max) for the durations of single activities of each type as well as for the numbers and sum of durations of all activities of each type on a day.

Table 4. The average, minimum, and maximum values for the durations (min) and numbers of activities.

	Duration for Single Activities			Number of All Activities on a Day			Sum of Durations of All Activities on a Day		
	Mean	**Min**	**Max**	**Mean ***	**Min**	**Max**	**Mean ***	**Min**	**Max**
S	312	60	720	2.22	1	4	693	418	1190
P	32	5	120	2.95	1	8	87	10	300
H	38	5	240	3.57	0	17	139	0	660
L	97	10	475	2.59	0	9	264	0	910
D	72	5	320	0.15	0	2	7	0	165
M	177	10	630	0.23	0	3	27	0	630

*: The mean is computed over all the schedules with or without activities of type k.

4.2. Prediction for Each Activity Record

Out of all the 36 explanatory variables (in Table 3), 34 for LM_{type} and 24, 27, 30, 25, 14, and 14 for the six sub-models of LM_{dur} (predicting the durations of S, P, H, L, D, and M, respectively) were significant (with p-value < 0.05), and they were selected as the predictors in the corresponding models. To measure the model performance, Acc_k and Acc_{all} are defined as the prediction accuracy by LM_{type} for activities of type k and of all types, respectively. Regarding LM_{dur}, the mean absolute percentage errors ($MAPE_{k,Z}$ and $MAPE_Z$), which are used to represent the average errors in the estimation of activity durations [38], are adopted. These variables can be computed as follows.

$$\mathrm{Acc_k} = \frac{REC_{\mathrm{k,cor}}}{REC_{\mathrm{k,all}}},\ \ \mathrm{Acc_{all}} = \frac{REC_{\mathrm{cor}}}{REC_{\mathrm{all}}} \tag{13}$$

$$\begin{aligned} MAPE_{k,Z} &= \frac{1}{REC_{k,all}} \sum_{i=1}^{REC_{k,all}} \left| \frac{\hat{Z}_{k,i} - Z_{k,i}}{Z_{k,i}} \right| \\ MAPE_Z &= \frac{1}{REC_{all}} \sum_{k=1}^{K} \left(\sum_{i=1}^{REC_{k,all}} \left| \frac{\hat{Z}_{k,i} - Z_{k,i}}{Z_{k,i}} \right| \right) \end{aligned} \tag{14}$$

In Equation (13), REC_{all} and REC_{cor} as well as $REC_{k,all}$ and $REC_{k,cor}$ represent the numbers of all and the correctly predicted activities of all types, as well as of type k, respectively, while in Equation (14), $\hat{Z}_{k,i}$ and $Z_{k,i}$ are the predicted and actual durations of activity i of type k.

Table 5 presents the results by LM_{type}, showing that S has the highest accuracy (Acc_k) of 0.8, followed by H, P, and L with 0.65, 0.66 and 0.63, while D and M suffer from the lowest accuracy with 0.56 and 0.58, respectively. The overall accuracy (Acc_{all}) is 0.67. In addition, it was noted that, among all the incorrect prediction results, H is the targeted type for most of the false positive prediction, with 7%, 22%, 21%, 19%, and 20% of S, P, L, D, and M being wrongly estimated as H. This can be attributed to two major reasons; H accounts for the largest share (i.e., 19%) among all the activity records, and the explanatory variables share common values between the records of H and those of the other types. Table 6 describes the results by LM_{dur}, and different levels of accuracy were also exposed. Specifically, S has the smallest error ($MAPE_{k,Z}$) of 0.14; P, H, and L have the median error of 0.18%, 0.17, and

0.15; while D and M exhibit the largest errors of 0.23 and 0.21, respectively. The overall error ($MAPE_Z$) is 0.16.

Table 5. Prediction results for activity types by LMtype.

Actual Types	Predicted Types					
	S	**P**	**H**	**L**	**D**	**M**
S	**0.80**	0.05	0.07	0.06	0.01	0.01
P	0.04	**0.65**	0.22	0.05	0.01	0.03
H	0.07	0.08	**0.66**	0.13	0.02	0.04
L	0.06	0.06	0.21	**0.63**	0.01	0.03
D	0.08	0.09	0.19	0.07	**0.56**	0.01
M	0.03	0.11	0.20	0.07	0.01	**0.58**

Table 6. Prediction results for activity durations by LM_{dur}.

	S	**P**	**H**	**L**	**D**	**M**	**All Types**
$MAPE_{k,Z}$	0.14	0.18	0.17	0.15	0.23	0.21	0.16

4.3. Prediction for Each Given Home Period

In *HomePPM*, Dur_{min} (i.e., the threshold for the minimum duration of activities) is designated as the minimum duration over all the in-home activities, i.e., 5 min. To measure the prediction performance, four variables are defined, including $MAE_{k,N}$ and $MAE_{k,U}$ as well as MAE_N and MAE_U. $MAE_{k,N}$ and $MAE_{k,U}$ represent the mean absolute errors for the estimation of the number of activities of type k and the sum of durations of these activities between each pair of the observed (i.e., *ScheO*) and predicted (i.e., *ScheI*) schedules, while MAE_N and MAE_U depict the average errors over all the types. These variables are computed as follows.

$$\begin{aligned}
MAE_{k,N} &= \frac{1}{\sum_{d=1}^{D} N_{k,d}} \sum_{d=1}^{D} \left|\hat{N}_{k,d} - N_{k,d}\right|, \quad MAE_{k,U} = \frac{1}{\sum_{d=1}^{D} N_{k,d}} \sum_{d=1}^{D} \left|\hat{U}_{k,d} - U_{k,d}\right| \\
MAE_N &= \frac{1}{\sum_{k=1}^{K}\sum_{d=1}^{D} N_{k,d}} \sum_{k=1}^{K} \left(\sum_{d=1}^{D} \left|\hat{N}_{k,d} - N_{k,d}\right|\right), \quad MAE_U = \frac{1}{\sum_{k=1}^{K}\sum_{d=1}^{D} N_{k,d}} \sum_{k=1}^{K} \left(\sum_{d=1}^{D} \left|\hat{U}_{k,d} - U_{k,d}\right|\right)
\end{aligned} \tag{15}$$

where, $N_{k,d}$ and $\hat{N}_{k,d}$ as well as $U_{k,d}$ and $\hat{U}_{k,d}$ are the observed and predicted numbers and sum of durations of activities of k in schedule d, respectively. Note that there are differences between $MAPE_{k,Z}$ and $MAE_{k,N}$ (or $MAE_{k,U}$). In $MAPE_{k,Z}$ (Equation (14)), the sum of the relative deviations between $\hat{Z}_{k,i}$ and $Z_{k,i}$ (i.e., $|\hat{Z}_{k,i} - Z_{k,i}|/Z_{k,i}$) is used, whereas in $MAE_{k,N}$ (or $MAE_{k,U}$) (Equation (15)), due to the possibility that $N_{k,d}$ and $U_{k,d}$ are equal to zero, the absolute (instead of relative) deviations between $\hat{N}_{k,d}$ and $N_{k,d}$ and between $\hat{U}_{k,d}$ and $U_{k,d}$ (i.e., $|\hat{N}_{k,d} - N_{k,d}|$ and $|\hat{U}_{k,d} - U_{k,d}|$) are considered.

The results obtained from *HomePPM* are listed in Table 7. Compared to the prediction accuracy (Acc_k and $MAPE_{k,Z}$) for single activity records, both similarities and differences were noted. Regarding $MAE_{k,N}$, S has the least error (i.e., 0.26), followed by P, H, and L (i.e., 0.37, 0.38, and 0.39), while D and M display the largest errors (i.e., 0.48 and 0.46), respectively. The error levels across the types are consistent with those for single activity records (i.e., 1-Acc_k = 0.2, 0.35, 0.34, 0.37, 0.44, and 0.42, respectively) (see Table 5). Nevertheless, due to the additional estimation for the number of activities in *HomeP*, $MAE_{k,N}$ has a higher level of errors than 1-Acc_k for each type, and the overall error MAE_N (0.36) is higher than 1-Acc_{all} (0.33). Similar phenomena occur for $MAE_{k,U}$, which has the smallest value for S (i.e., 0.18), middle values for P, H, and L (i.e., 0.21, 0.22, and 0.24), and largest values for D and M (i.e., 0.28 and 0.26), in line with the error trend of $MAPE_{k,Z}$ for single activity records. Moreover, $MAE_{k,U}$ is larger than $MAPE_{k,Z}$ for each type (see Table 7), and the overall error MAE_U (0.21) is higher than $MAPE_Z$ (0.16). In addition to the forecasted number of activities, the increased errors of $MAE_{k,U}$ can be attributed to the process of *HomePPM,* in which the prediction of activity durations is based on the previously derived activity types and thus is affected by the accuracy of the estimated types.

Table 7. Prediction results by *HomePPM* and by the ASI-based method.

		S	P	H	L	D	M	All Types
HomePPM	$MAE_{k,N}$	0.26	0.37	0.38	0.39	0.48	0.46	0.36
	$MAE_{k,U}$	0.18	0.21	0.22	0.24	0.28	0.26	0.21
ASI-based method	$MAE_{k,N}$	0.22	0.25	0.26	0.29	0.32	0.30	0.26
	$MAE_{k,U}$	0.11	0.13	0.14	0.14	0.16	0.14	0.13
Differences	$MAE_{k,N}$	0.04	0.12	0.12	0.10	0.16	0.16	0.10
	$MAE_{k,U}$	0.07	0.08	0.08	0.10	0.12	0.12	0.08

4.4. ASI-Based Enhancement

Tables 8 and 9 depict $Tr(k_2 \mid k_1)$ and $Qr(k_2 \mid k_1)$; large differences were noted between these two types of factors. In Table 9, the highest transition factors from in-home activities are dominated by the transitions to H, P, and L. In contrast, in Table 9, the dominance of these types is reduced by their high frequencies, and transitions to other less-common activities (e.g., D and M) are exposed. Furthermore, as reflected in Table 9, there exists a certain degree of correlation between in-home and out-of-home activities, represented by the varied values of in-home (or out-of-home) activities after an out-of-home (or in-home) activity. For example, after personal care (P) at home, the most likely out-of-home activities are D_o and M_o (i.e., 'P-D_o' and 'P-M_o'); while after mandatory activities (M_o) outside home, the most oriented in-home activities are M and P (i.e., 'M_o-M and 'M_o-P). In addition, a relationship was also observed across different types of in-home activities, e.g., 'S-P', 'P-L' and 'L-S'. Particularly, the same types of activities are likely to be chained together, e.g., 'H-H', 'D-D' and 'M-M'. All the above results further confirm the statements [7,18] that the characteristics (e.g., activity types and durations) of activities conducted prior to and directly following an activity have a significant impact on the location choice and characteristics of the activity.

Table 8. Transition matrix $Tr(k_2 \mid k_1)$ *.

Previous Activity	Succeeding Activity										
	S	P	H	L	D	M	P_o	H_o	L_o	D_o	M_o	SH_o
S	0.18	**0.39**	0.23	0.10	0.01	0.01	0.01	0.02	0.02	0.004	0.01	0.01
P	0.10	0.13	0.26	**0.29**	0.01	0.02	0.02	0.03	0.04	0.02	0.05	0.03
H	0.04	0.28	**0.35**	0.21	0.01	0.02	0.01	0.03	0.02	0.01	0.01	0.02
L	**0.24**	0.21	**0.24**	0.18	0.01	0.02	0.01	0.02	0.03	0.01	0.01	0.03
D	0.14	0.19	**0.24**	0.22	0.06	0.02	0.01	0.02	0.02	0.03	0.01	0.04
M	0.08	0.22	**0.24**	0.23	0.02	0.06	0.02	0.04	0.03	0.01	0.04	0.03
P_o	0.02	0.05	0.06	0.08	0.003	0.01	0.04	0.09	**0.23**	0.02	0.30	0.09
H_o	0.02	0.09	0.19	0.08	0.003	0.01	0.08	**0.20**	0.12	0.03	0.06	0.10
L_o	0.05	0.13	0.11	0.09	0.003	0.01	0.11	0.09	**0.17**	0.04	0.13	0.07

Table 8. *Cont.*

Previous Activity	Succeeding Activity											
	S	**P**	**H**	**L**	**D**	**M**	**P_o**	**H_o**	**L_o**	**D_o**	**M_o**	**SH_o**
D_o	0.01	0.11	0.12	0.07	0.01	0.01	0.09	0.06	0.10	**0.26**	0.02	0.14
M_o	0.02	0.12	0.10	0.05	0.002	0.01	**0.25**	0.06	0.13	0.02	0.16	0.07
SH_o	0.01	0.11	**0.27**	0.09	0.002	0.01	0.11	0.07	0.07	0.02	0.03	0.20

* The largest value in each row is in bold, and the sum over all values of each row is 1.

Table 9. Transition matrix $Qr(k_2 \mid k_1)$ *.

Previous Activity	Succeeding Activity											
	S	**P**	**H**	**L**	**D**	**M**	**P_o**	**H_o**	**L_o**	**D_o**	**M_o**	**SH_o**
S	13.8	**14.4**	6.8	4.0	7.0	5.9	2.6	3.9	2.5	1.5	1.7	1.7
P	7.6	4.6	7.8	**12.2**	9.6	10.8	3.0	4.6	5.4	8.1	7.8	4.5
H	2.7	10.1	**10.5**	8.7	7.6	7.2	1.1	5.5	2.4	3.3	2.0	3.7
L	**18.0**	7.8	7.1	7.5	9.3	7.4	2.1	3.4	3.7	3.7	1.5	4.4
D	10.5	7.1	7.2	9.1	**44.7**	9.7	1.3	2.7	2.9	11.1	1.9	6.4
M	6.1	8.2	7.2	9.3	13.4	**28.1**	3.0	5.8	4.2	2.6	5.9	4.2
P_o	1.9	1.8	1.9	3.4	1.8	3.0	7.7	14.2	31.5	6.5	**46.1**	14.9
H_o	1.7	3.5	5.6	3.4	2.4	6.0	16.2	**32.3**	15.9	10.1	9.6	16.2
L_o	3.7	4.7	3.4	3.8	2.9	3.1	20.4	14.2	**22.6**	12.3	20.4	12.1
D_o	9.9	3.9	3.7	3.0	4.1	2.8	17.6	9.3	13.4	**90.2**	3.8	23.1
M_o	1.8	4.6	3.0	2.2	1.6	4.7	**48.5**	10.0	17.1	6.4	23.7	12.0
SH_o	8.5	4.0	8.0	3.7	1.8	3.4	22.1	11.5	10.0	8.7	5.1	**32.1**

*: The largest value in each row is in bold, and the actual probability for each transition is the cell value multiplied by 1×10^{-7}.

After testing on different values, TH_1 and TH_2 were set as 0.5 and 0.9 respectively; the obtained prediction results are presented in Table 8. Compared to those before the enhancement, $MAE_{k,N}$ decreases by 0.04, 0.12, 0.12, 0.10, 0.16, and 0.16 for S, P, H, L, D, and M, with an overall decrease of 0.10. This reduction also leverages the estimate for durations, leading to $MAPE_{k,U}$ falling by 0.07, 0.08, 0.08, 0.10, 0.12, and 0.12, with an overall decrease of 0.08. When the reduced errors were compared across different activity types, it was noted that, while the ASI-based method strengthens the prediction for all the types, this method particularly enhances the accuracy for less-common activity types (e.g., D and M). This can be due to the fact that the statistical models (or machine learning algorithms) usually favor majority classes if the classification accuracy is used as the model evaluation criterion, whereas the ASI-based method puts equal weights on all classes of the dependent variable (i.e., the in-home activity types).

4.5. Model Calibration

To inspect the practical ability of the MA calibration process, the ABM schedules generated from a synthetic population of Flanders, Belgium by a representative ABM, Feathers, were used. We acknowledged that the real activity behavior between Belgium and the United States is likely to be different on a certain degree. Consequently, the illustration serves to underline the applicability of the calibration, but not to infer activity behavioral relationships between these two countries. The analysis was carried out in two steps, including the prediction of in-home activities using the MB process, and the improvement of the predicted results by the MA calibration.

4.5.1. Feathers

Feathers [39] is an ABM that provides detailed spatial-temporal microsimulations for human mobility. It allows for more realistic and consistent linkages across activity and travel choices made by individuals in the course of a day, simulating individual agents along with their prediction of activity–travel schedules [40]. Based on Feathers, the schedule of each individual (older than 6) in the Flemish region of Belgium was generated, resulting in a total of 6.3 M (million) daily sequences, accounting for 94% of the population. The ABM schedules accommodate 14.2 M home periods, with 2.25 periods per person. The average time spent at home is 1149 min per schedule (79.8% of the day) and 510 min per period.

4.5.2. The Calibration Process

Based on the MB process, the in-home activities of the ABM schedules were predicted, forming a set of initially estimated schedules (i.e., *ScheFs*). *ScheFs* were further improved based on the calibration process, resulting in the finally obtained schedules (i.e., *ScheCs*). While the observed marginal activity data can be obtained from various sources, e.g., the survey on economics and life quality; in this study, the information was collected from the Belgian official site of statistics on the Belgian economy, society, and territory [41]. Moreover, without losing generalization, two explanatory variables were considered, including the gender (Sex) and employment status (Work) of respondents. During the calibration, five parameters (including A, s, η, b, and γ) were defined, and these parameters control the optimization process through s_n and b_n. A small value of s_n or b_n may lead to the algorithm being stuck in the current position regardless of whether it is optimal or not; whereas a large value could cause the algorithm to take a big step far away from the optimal solution. A more detailed discussion on these parameters can be referred to in [35]. In this study, the values suggested by [36] were adopted, which are 100, 0.2, 0.602, 0.5, and 0.101 for A, s, η, b, and γ, respectively. For performance measures, $MAPE_{type}$ and $MAPE_{dur}$ are used to represent the average percentage differences between the observed and predicted participation rates and time spent over all the MA variables, respectively. They are computed as follows:

$$\begin{aligned} MAPE_{type} &= \frac{1}{Var}\left(\sum_{k=1}^{K}\left|\frac{\hat{R}_k - R_k}{R_k}\right| + \sum_{j=1}^{m}\sum_{c=1}^{C_m}\sum_{k=1}^{K}\left|\frac{\hat{R}_{k,m,c} - R_{k,m,c}}{R_{k,m,c}}\right|\right) \\ MAPE_{dur} &= \frac{1}{Var}\left(\sum_{k=1}^{K}\left|\frac{\hat{T}_k - T_k}{T_k}\right| + \sum_{j=1}^{m}\sum_{c=1}^{C_m}\sum_{k=1}^{K}\left|\frac{\hat{T}_{k,m,c} - T_{k,m,c}}{T_{k,m,c}}\right|\right) \end{aligned} \quad (16)$$

where, K = 6, m = 2, C_1 = 2 for Sex (including men and women) and C_2 = 3 for Work (including unemployed, full-time, and part-time), leading to the total number of the MA variables (i.e., *Var*) for participation rates (R_k and $R_{k,m,c}$) and time spent (T_k and $T_{k,m,c}$) to be 36.

4.5.3. Calibration Results

Figure 5a,b describes the evolution of the object functions $O_{type}(\beta)$ and $O_{dur}(\alpha)$ when the number of iterations (n) increases, showing that $O_{type}(\beta)$ and $O_{dur}(\alpha)$ gradually decrease and reach the lowest points around n = 500 and n = 100, respectively. To obtain the minimum values of both object functions, TH_{ite} was specified as 500, under which the differences between $O_{type}(\beta_{n+1})$ and $O_{type}(\beta_n)$ as well as between $O_{dur}(\alpha_{n+1})$ and $O_{dur}(\alpha_n)$ over several successive iterations are smaller than 0.0001.

Table 10 describes the values of R_k, $\hat{R}_{k1}$, and $\hat{R}_{k2}$ as well as T_k,$\hat{T}_{k1}$, and $\hat{T}_{k2}$, where $\hat{R}_{k1}$ and $\hat{T}_{k1}$ as well as $\hat{R}_{k2}$ and $\hat{T}_{k2}$ are extracted from ScheFs and ScheCs over all individuals, respectively, and $\Delta R_{k1} = \hat{R}_{k1} - R_k$, $\Delta R_{k2} = \hat{R}_{k2} - R_k$, $\Delta T_{k1} = \hat{T}_{k1} - T_k$, and $\Delta T_{k2} = \hat{T}_{k2} - T_k$. It shows that, for each activity type, the absolute values of ΔR_{k1} and ΔT_{k1} (before the calibration) are larger than those of ΔR_{k2} and ΔT_{k2} (after the calibration) (i.e., $|\Delta R_{k1}| > |\Delta R_{k2}|$ and $|\Delta T_{k1}| > |\Delta T_{k2}|$). Similar results were derived for the other MA variables (i.e., $R_{k,m,c}$, $\hat{R}_{k,m,c}$, $T_{k,m,c}$, and $\hat{T}_{k,m,c}$). Over all the MA variables, $MAPE_{type}$ and $MAPE_{dur}$ were obtained as 0.35 and 0.28 from ScheFs and 0.14 and 0.09 from ScheCs; an overall improvement of 60% and 68% for participation rates and time spent was gained by the calibration.

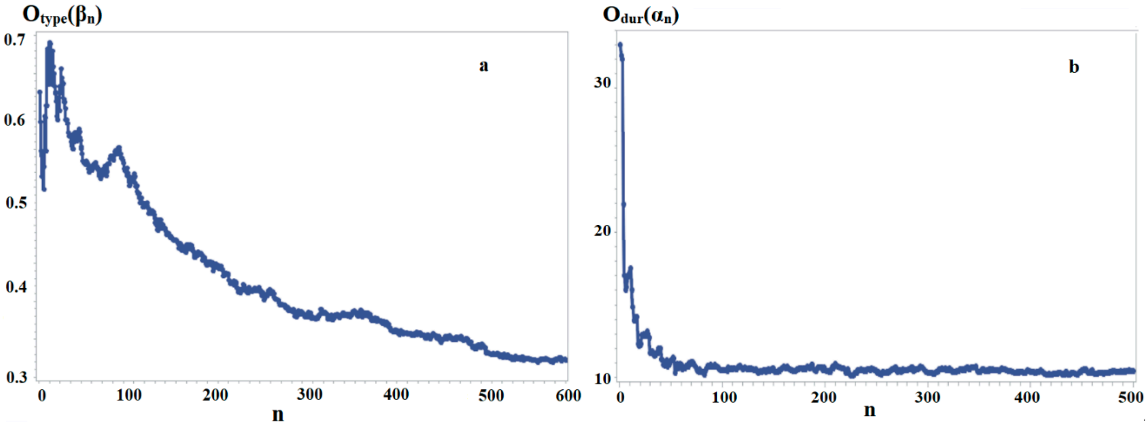

Figure 5. The values of $O_{type}(\beta)$ (**a**) and $O_{dur}(\alpha)$ (**b**) over the iterations.

Table 10. The predicted and observed participation rates and time spent for all individuals.

	Participation Rates					**Time Spent (min)**				
	Observed	***ScheFs***		***ScheCs***		**Observed**	***ScheFs***		***ScheCs***	
	R_k	$\hat{R}_{k1}$	ΔR_{k1}	$\hat{R}_{k2}$	ΔR_{k2}	T_k	$\hat{T}_{k1}$	ΔT_{k1}	$\hat{T}_{k2}$	ΔT_{k2}
S	1	0.98	−0.02	0.98	−0.02	543	579	36	548	5
P	0.96	0.75	−0.21	0.85	−0.11	147	121	−26	134	−13
H	0.86	0.71	−0.15	0.75	−0.11	152	222	70	166	14
L	0.95	0.64	−0.31	0.84	−0.11	242	194	−48	239	−3
D	0.12	0.03	−0.09	0.09	−0.03	5	3	−2	4	−1
M	0.08	0.03	−0.05	0.06	−0.02	34	21	−13	32	−2

4.5.4. Comparing Between ScheFs and ScheCs

To further examine the differences between ScheFs and ScheCs, the average numbers of activities of k per schedule (i.e., N_k, $\hat{N}_k$, $\hat{N}_{k1}$, and $\hat{N}_{k2}$) over all ScheOs, ScheIs, ScheFs, and ScheCs were derived. In addition, using N_k as a reference, the differences between N_k and the other variables were computed (i.e., $\Delta N_k = \hat{N}_k - N_k$, $\Delta N_{k1} = \hat{N}_{k1} - N_k$, and $\Delta N_{k2} = \hat{N}_{k2} - N_k$). These results are presented in Table 11.

Table 11. Average number of activities of each type per schedule.

	ScheOs	***ScheIs***			***ScheFs***			***ScheCs***		
	N_k	$\hat{N}_k$	ΔN_k	$\Delta N_k/N_k$	$\hat{N}_{k1}$	ΔN_{k1}	$\Delta N_{k1}/N_k$	$\hat{N}_{k2}$	ΔN_{k2}	$\Delta N_{k2}/N_k$
S	2.22	2.54	0.32	0.14	2.48	0.26	0.12	2.30	0.08	0.04
P	2.95	2.50	−0.45	−0.15	2.16	−0.79	−0.27	2.61	−0.34	−0.11
H	3.57	4.43	0.86	0.24	5.50	1.93	0.54	4.39	0.82	0.23
L	2.59	2.09	−0.5	−0.19	1.66	−0.93	−0.36	2.12	−0.47	−0.18
D	0.15	0.10	−0.05	−0.33	0.08	−0.07	−0.47	0.11	−0.04	−0.27
M	0.23	0.16	−0.07	−0.30	0.13	−0.10	−0.43	0.18	−0.05	−0.22

Based on Tables 10 and 11, the following important features were noticed. (1) The MB process tends to over-estimate certain activity types (e.g., S and H) while it under-estimate other types (e.g., D and M). This can be demonstrated by the results in Table 11, where $\hat{N}_k$ and $\hat{N}_{k1}$ increase by 14% and 12% for S and 24% and 54% for H, whereas they decrease by 33% and 47% for D and 30% and 43% for M, respectively, when compared to the observed number (N_k). (2) However, despite the over- or under-estimation of individual activity types, the participate rate $\hat{R}_{k1}$ is lower than the actually observed rate R_k, leading to $\Delta R_{k1} \leq 0$ for all the types (see Table 10). This suggests that, most of the activities for a same schedule tend to be forecasted with identical types, which could be due to the fact that for a same person, most of the explanatory variables (e.g., the individual–household variables) are unchanged, and only a few (e.g., the activity–schedule variables) may differ. Consequently, each *ScheF* (or *ScheI*) is inclined to contain more activities of the same types, whereas it lacks a variety of different types within the daily sequence. This leads to the percentage of individuals who perform at least one activity of a given type being low. (3) Affected by the biased estimation for individual activity types, the time spent $\hat{T}_{k1}$ also shows a certain degree of orientation, with $\hat{T}_{k1}$ increasing for S and H (i.e., $\Delta T_{k1} > 0$) but reducing for D and M (i.e., $\Delta T_{k1} < 0$) in relation to the observed time (T_k) (see Table 10).

In comparison, the calibration improves the initial prediction results by adjusting the model parameters (in LM_{type} and LM_{dur}) in order to reduce the discrepancies between the predicted and observed marginal activity variables. During this process, the above, biased estimations were mitigated, as manifested by the following results. (1) The occurrences of over-estimated activity types (e.g., S and H) decreased, while those of the under-estimated types (e.g., D and M) increased, resulting in the deviations between N_k and $\hat{N}_{k2}$ being smaller (than between N_k and $\hat{N}_{k1}$). For instance, $\hat{N}_{k2}$ increases by 4% and 23% for S and H while it decreases by 27% and 22% for D and M relative to N_k (see Table 11), demonstrating smaller changes than $\hat{N}_{k1}$ does for the same types (i.e., $|\Delta N_{k2}/N_k| < |\Delta N_{k1}/N_k|$). (2) The diversity of activity types per schedule was enlarged, leading to $\hat{R}_{k2}$ elevating and the difference between $\hat{R}_{k2}$ and R_k diminishing (i.e., $|\Delta R_{k2}| < |\Delta R_{k1}|$) (see Table 11). (3) The variations in time spent between the observed and predicted schedules were also reduced, with $|\Delta T_{k2}|$ for each type being smaller than its counterpart $|\Delta T_{k1}|$ (see Table 10).

5. Discussion

In this study, a new method of enriching in-home activities for ABM schedules was proposed, based on the integration of statistical modeling, activity sequential information and calibration. Given a number of well-established ABMs across the world, the proposed method can be generically applied to the output of these models and can contribute to the extension of ABMs to a wide range of applications that are associated with individuals' in-home activities (e.g., energy consumption and carbon emission estimations).

By applying the approach to real activity–travel schedules in the test set as well as to the millions of ABM-schedules generated by Feathers, the potential and practical ability of the method were evaluated. In an ideal situation, we would have used data from Belgium time use surveys for model estimation; this may negate the advantage in terms of improving the predictions using the calibration step, but this step has been recommended in the transport literature in order to match the sample surveyed aggregate outcomes with population-based aggregate outcomes [42]. Because of the unavailability of the raw time survey from Belgium, we used the US time use survey for model estimation and then performed a calibration process to match the marginal distributions (i.e., aggregate statistics from the Belgium time use survey). This workflow has an advantage and provides a holistic methodological approach, which is suitable for the availability of datasets in different contexts (as is the case here) and provides flexibility and transferability benefits, but the calibration step is equally important, even if the datasets are consistent in terms of time and space. Detailed surveys often represent a small sample of the population, and models estimated using these surveys require a calibration process to match population statistics. Based on the statistical modeling (*HomePPM*), the mean absolute errors (*MAEN* and *MAEU*) were 0.36 and 0.21 for the prediction of the number and sum of durations of in-home activities (over all types) per schedule, respectively. By means of the ASI-based approach, the prediction obtained a 10% and 8% improvement. After the MA calibration process, an additional advancement of 60% and 68% (*MAPEtype* and *MAPEdur*) was gained regarding the activity participation rates and time spent per day.

Individual–household and activity attributes both played vital roles in appropriately predicting the in-home activity types and their duration. This is in line with the findings of the existing literature, where statistical models were used to model these two outcomes [17,18,20]. In total, we had 36 such variables and in all estimated models, the majority of them were found to be significant. Additionally, the ASI based enhancement proved its importance in further adjusting and ordering the right activities in an appropriate sequence. The ASI-based enhancement is based on the fact that certain activities have more chances to appear next in the schedule based on what activities were conducted before. The results obtained confirmed the findings of [7,18].

Nevertheless, despite the promising results, there is still a certain level of misclassifications. To further reduce the errors, enhancement could be performed in each major step of the approach. (1) In *LMtype* and *LMdur*, more comprehensive models (e.g., mixed logit models and hazard functions) could be adopted, and a non-linear correlation between the dependent and explanatory variables could be considered [14,16]. In addition, information on a broad picture of individuals' daily schedules and habits (e.g., how often an individual works or shops online at home) could be used as extra explanatory variables. (2) Regarding the *HomePPM* method, an additional process could be utilized to examine the total number and duration of in-home activities of each type k over all the home periods of a schedule, in order to ensure that the number and duration fall between corresponding minimum and maximum values for k of an observed schedule (see Table 4). Through the above checking, the possibly biased estimation of activity types in a schedule (as described in Section 4.4) could be alleviated. (3) In terms of the ASI-based method, the transition probabilities Qr(k2 | k1) should be derived for different time periods of the day (e.g., morning, afternoon, and evening), as activity sequential patterns may differ across these periods. (4) With respect to calibration, more MA variables can be utilized, and the threshold THite can be set larger (than the present value), which, however, requires more iterations and a longer running time.

As indicated in the literature review, there are existing methods to forecast in-home activities, for instance, a method that builds the prediction process within an ABM modeling framework (ADAPTS) [6,7] (See Section 2.3). However, such ABMs are quite scarce, and it is not easy to overhaul an operational ABM to include in-home activities in its core modeling framework. The proposed approach provides a more flexible way to enrich the outputs of operational ABMs without interfering with the core modeling components. Therefore, the ASI-based method and MA calibration process proposed in this study, along with the previously suggested methods for major enhancement, can be easily integrated into the existing ABM approaches in order to advance the prediction methods that do not only provide high accuracy (in relation to the model training and testing data) but also a good match with activity sequential patterns and actual marginal activity distributions.

6. Conclusions

This study proposes a new method to predict the types and durations of in-home activities using activity–travel schedules from an activity-based travel demand model (ABM). The method uses statistical methods like multinomial logit, log-linear regression, and activity sequential information and calibration process based on the SPSA algorithm. Tested on 6.3 million people in Belgium, the method showed a 10% and 8% improvement in prediction accuracy using sequential information. After calibration, it gained an additional 60% and 68% in activity participation rates and time spent per day. With a few limitations, the method has the potential to incorporate in-home activities into the outcomes of ABMs for various applications, such as evaluating energy consumption and carbon emissions in different sustainable urban policy contexts.

Author Contributions: Conceptualization, M.A. and F.L.; methodology, M.A. and F.L.; investigation, D.J., T.B., M.A. and F.L.; resources, M.A.; data curation, F.L.; writing—original draft preparation, M.A. and F.L.; writing—review and editing, D.J. and T.B.; supervision, M.A. and G.W.; project administration, M.A.; funding acquisition, M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is based on a project that has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 101037193.

Institutional Review Board Statement: This study did not requrie ethical approval.

Informed Consent Statement: This study used secondary source of data availabel open source (and provided reference as well), the study does not involve human.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author/s. The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the study's design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Definition of certain major variables.

Variables	Definition
Activities	
K and k	The total number of in-home activity types (K) and each of these types (k) (k = 1, . . ., K)
M, x_m and C_m	The total number of explanatory variables (M), each of these variables (x_m) (m = 1, . . ., M), and the number of categories of x_m (C_m)
Y and Z	The dependent variables for in-home activity types (Y) and activity duration (Z)
$\hat{Z}_{k,i}$ and $Z_{k,i}$	The predicted ($\hat{Z}_{k,i}$) and observed ($Z_{k,i}$) durations of activity i of type k
$MAPE_{k,Z}$ and $MAPE_Z$	The mean absolute percentage error in the estimation of durations for activities of type k and of all types ($MAPE_Z$) (see Equation (14))
Schedules	
ABM-schedule and *HomeP*	A daily schedule generated by ABMs (ABM schedule), and each given home period of the schedule (*HomeP*) that is to be enriched with in-home activities
ScheO, *ScheI*, *ScheF* and *ScheC*	An observed (*ScheO*) and corresponding predicted (i.e., in-home activity enriched) (*ScheI*) schedule from *ATUS-wd*; the predicted schedule for an ABM-schedule of Feathers before (*ScheF*) and after (*ScheC*) calibration
d and D	Each schedule (d) and the total number of schedules (D) in the relevant set
$N_{k,d}$,$\hat{N}_{k,d}$, $U_{k,d}$ and $\hat{U}_{k,d}$	The observed ($N_{k,d}$) and predicted ($\hat{N}_{k,d}$) numbers of activities of k in d; the observed ($U_{k,d}$) and predicted ($\hat{U}_{k,d}$) sum of durations of activities of k in d
$MAE_{k,N}$, $MAE_{k,U}$, MAE_N and MAE_U	The mean absolute error in the estimation of the number ($MAE_{k,N}$) and sum of durations ($MAE_{k,U}$) of activities of k between each pair of the observed and predicted schedules; the average of $MAE_{k,N}$ (MAE_N) and $MAE_{k,U}$ (MAE_U) over all the types (see Equation (15))
Activity sequential information	
$Tr(k_2 \mid k_1)$ and $Qr(k_2 \mid k_1)$	Transition probabilities from activity types k_1 to k_2 (k_1, k_2 = 1, . . ., K)
$Pr^T(Y = k_2)$ and $Pr^Q(Y\ k_2)$	New probabilities of activities with the type of k_2 given the previous activity of k_1

Table A1. *Cont.*

Variables	Definition
	Marginal activity variables
R_k, $\hat{R}_k$, T_k and $\hat{T}_k$	The observed (R_k) and predicted percentage ($\hat{R}_k$) of individuals who perform at least one activity of k each day (schedule), and the observed (T_k) and predicted ($\hat{T}_k$) average time spent on activities of k on a day (schedule) per person (see Equation (11))
$R_{k,m,c}$, $\hat{R}_{k,m,c}$, $T_{k,m,c}$ and $\hat{T}_{k,m,c}$	The observed ($R_{k,m,c}$) and predicted ($\hat{R}_{k,m,c}$) percentage of people with $x_m = c$ who perform activities of k, and the observed ($T_{k,m,c}$) and predicted ($\hat{T}_{k,m,c}$) average time spent on activities of k per person per day among this group.
$MAPE_{type}$ and $MAPE_{dur}$	The average absolute percentage difference between the observed and predicted participation rates ($MAPE_{type}$) and time spent ($MAPE_{dur}$) over all the activity types (see Equation (16))

References

1. Damm, A.; Köberl, J.; Prettenthaler, F.; Rogler, N.; Töglhofer, C. Impact of +2 °C global warming on electricity demand in Europe. *Clim. Serv.* **2017**, *7*, 12–30. [CrossRef]
2. Watts, N.; Amann, M.; Arnell, N.; Ayeb-Kalsson, S.; Belesova, K. The 2019 report of The Lancet Countdown on health and climate change: Ensuring that health of a child born today is not defined by a changing climate. *Lancet* **2019**, *394*, 1836–1878. [CrossRef] [PubMed]
3. I-CHANGE. 2022. Available online: https://ichange-project.eu/ (accessed on 15 November 2024).
4. Arentze, T.A.; Timmermans, H.J.P. A learning-based transportation oriented simulation system. *Transp. Res. Part B Methodol.* **2004**, *38*, 613–633. [CrossRef]
5. Wegener, M. The Future of Mobility in Cities: Challenges for Urban Modelling. *Transp. Policy* **2013**, *29*, 275–282. [CrossRef]
6. Langerudi, M.F.; Javanmardi, M.; Shabanpour, R.; Rashidi, T.H.A. Incorporating in-home activities in ADAPTS activity-based framework: A sequential conditional probability approach. *J. Transp. Geogr.* **2017**, *61*, 48–60. [CrossRef]
7. Shabanpour, R.; Golshani, N.; Langerudi, M.F.; Mohammadian, A.K. Planning in-home activities in the ADAPTS activity-based model: A joint model of activity type and duration. *Int. J. Urban Sci.* **2018**, *22*, 236–254. [CrossRef]
8. U.S. Bureau of Labour Statistics. 2019. Available online: https://www.bls.gov/tus/database.html (accessed on 15 November 2024).
9. Kahanec, M.; Lafférs, L.; Marcus, J.S. The Impact of COVID-19 Restrictions on Individual Mobility. Bruegel Blog, 5 May 2020. Available online: https://www.bruegel.org/2020/05/the-impact-of-covid-19-restrictions-on-individual-mobility (accessed on 15 November 2024).
10. Cole, M.A.; Ozgen, C.; Strobl, E. Air Pollution Exposure and COVID-19. Institute of Labor Economics (IZA), IZA Discussion Papers 13367. 2020. Available online: https://ideas.repec.org/p/iza/izadps/dp13367.html (accessed on 15 November 2024).
11. Mitra, D.; Chu, Y.Y.; Cetin, K. COVID-19 impacts on residential occupancy schedules and activities in US Homes in 2020 using ATUS. *Appl. Energy* **2022**, *324*, 119765. [CrossRef]
12. Hagerstrand, T. What about people in regional science? In *Papers of the Regional Science Association*; Springer: Berlin/Heidelberg, Germany, 1970; pp. 7–21.
13. Jones, P.M. New approaches to understanding travel behaviour: The human activity approach. In *Behavioural Travel Modelling*; Routledge: London, UK, 2021; pp. 55–80.
14. Davidson, W.; Donnelly, R.; Vovsha, P.; Freedman, J.; Ruegg, S.; Hicks, J.; Castiglione, J.; Picado, R. Synthesis of first practices and operational research approaches in activity-based travel demand modeling. *Transp. Res. Part A Policy Pract.* **2007**, *41*, 464–488. [CrossRef]
15. Akar, G.; Clifton, K.J.; Doherty, S.T. Discretionary activity location choice: In-home or out-of-home? *Transportation* **2011**, *38*, 101–122. [CrossRef]
16. Bhat, C.R.; Koppelman, F.S. A Retrospective and Prospective Survey of Time-Use Research. *Transportation* **1999**, *26*, 119–139. [CrossRef]
17. Yamamoto, T.; Kitamura, R. An analysis of time allocation to in-home and out-of-home discretionary activities across working days and non-working days. *Transportation* **1999**, *26*, 231–250. [CrossRef]
18. Bhat, C.R.; Gossen, R. A mixed multinomial logit model analysis of weekend recreational episode type choice. *Transp. Res. Part B* **2004**, *38*, 767–787. [CrossRef]
19. Doherty, S.T. Should We Abandon Activity Type Analysis? Redefining Activities by Their Salient Attributes. *Transportation* **2006**, *33*, 517–536. [CrossRef]
20. Miller, E.; Roorda, M.J. A Prototype Model of Household Activity/Travel Scheduling. *J. Transp. Res. Board* **2003**, *1831*, 114–121. [CrossRef]

21. Javadinasr, M.; Maggasy, T.; Mohammadi, M.; Mohammadain, K.; Rahimi, E.; Salon, D.; Conway, M.W.; Pendyala, R.; Derrible, S. The Long-Term effects of COVID-19 on travel behavior in the United States: A panel study on work from home, mode choice, online shopping, and air travel. *Transp. Res. Part F Psychol. Behav.* **2022**, *90*, 466–484. [CrossRef]
22. Huang, Z.R.; Loo, B.P.Y.; Axhausen, K.W. Travel behaviour changes under Work-from-home (WFH) arrangements during COVID-19. *Travel Behav. Soc.* **2023**, *30*, 202–211. [CrossRef]
23. Raišienė, A.G.; Rapuano, V.; Varkulevičiūtė, K.; Stachová, K. Working from Home—Who Is Happy? A Survey of Lithuania's Employees during the COVID-19 Quarantine Period. *Sustainability* **2020**, *12*, 5332. [CrossRef]
24. Barbour, N.; Menon, N.; Mannering, F. A statistical assessment of work-from-home participation during different stages of the COVID-19 pandemic. *Transp. Res. Interdiscip. Perspect.* **2021**, *11*, 100441. [CrossRef]
25. Dubey, A.D.; Tripathi, S. Analysing the Sentiments towards Work-From-Home Experience during COVID-19 Pandemic. *J. Innov. Manag.* **2020**, *8*, 13–19. [CrossRef]
26. Salon, D.; Mirtich, L.; Bhagat-Conway, M.W.; Costello, A.; Rahimi, E.; Mohammadian, A.K.; Chauhan, R.S.; Derrible, S.; da Silva Baker, D.; Pendyala, R.M. The COVID-19 pandemic and the future of telecommuting in the United States. *Transp. Res. Part D Transp. Environ.* **2022**, *112*, 103473. [CrossRef]
27. Elldér, E. Telework and daily travel: New evidence from Sweden. *J. Transp. Geogr.* **2020**, *86*, 102777. [CrossRef] [PubMed]
28. Shamshiripour, A.; Rahimi, E.; Shabanpour, R.; Mohammadian, A.K. How is COVID-19 reshaping activity-travel behavior? Evidence from a comprehensive survey in Chicago. *Transp. Res. Interdiscip. Perspect.* **2020**, *7*, 100216. [CrossRef] [PubMed]
29. Khalil, M.A.; Fatmi, M.R. Modeling daily in-home activities using machine learning techniques. *Travel Behav. Soc.* **2023**, *31*, 374–385. [CrossRef]
30. Hesam Hafezi, M.; Sultana Daisy, N.; Millward, H.; Liu, L. Framework for development of the Scheduler for Activities, Locations, and Travel (SALT) model. *Transp. A Transp. Sci.* **2021**, *18*, 248–280. [CrossRef]
31. Liu, F.; Janssens, D.; Cui, J.X.; Wets, G.; Cools, M. Characterizing activity sequences using Profile Hidden Markov Models. *Expert Syst. Appl.* **2015**, *42*, 5705–5722. [CrossRef]
32. United States Census Bureau. 2022. Available online: https://www.census.gov/about.html (accessed on 15 November 2024).
33. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann: San Mateo, CA, USA, 1993.
34. Zheng, Y.; Chen, Y.K.; Xie, X.; Ma, W.Y. Understanding transportation modes based on GPS data for Web applications. *ACM Trans. Web Assoc. Comput. Mach.* **2010**, *4*, 1–36. [CrossRef]
35. Spall, J.C. Multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Trans. Autom. Control.* **1992**, *37*, 332–341. [CrossRef]
36. Flötteröd, G. A search acceleration method for optimization problems with transport simulation constraints. *Transp. Res. Part B Methodol.* **2017**, *98*, 239–260. [CrossRef]
37. Lu, L.; Xu, Y.; Antoniou, C.; Ben-Akiva, M.A. An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models. *Transp. Res. Part C Emerg. Technol.* **2015**, *51*, 149–166. [CrossRef]
38. Washington, S.; Karlaftis, M.; Mannering, F. *Statistical and Econometric Methods for Transportation Data Analysis*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2011.
39. FEATHERS. 2017. Available online: https://www.abeonaconsult.be/feathers/ (accessed on 15 November 2024).
40. Hoang, T.L.; Adnan, M.; Vu, A.T.; Hoang-Tung, N.; Kochan, B.; Bellemans, T. Modeling and Structuring of Activity Scheduling Choices with Consideration of Intrazonal Tours: A Case Study of Motorcycle-Based Cities. *Sustainability* **2022**, *14*, 6367. [CrossRef]
41. Statbel. 2017. Available online: https://statbel.fgov.be/en/themes/households/time-use-survey#figures (accessed on 15 November 2024).
42. Ortuzar, J.D.; Willumsen, L.G. *Modelling Transport*, 4th ed.; John Wiley and Sons: Hoboken, NJ, USA, 2011.