Energy & Buildings 326 (2025) 115036

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enbuild

ETHOS.ActivityAssure—An open-source validation framework for synthetic European activity profiles

David Neuroth [a,b,*], Noah Pflugradt [a], Jann Michael Weinand [a], Christina Büsing [b], Detlef Stolten [a,c]

[a] Forschungszentrum Jülich GmbH, Institute of Climate and Energy Systems – Jülich Systems Analysis (ICE-2), 52425, Jülich, Germany
[b] Combinatorial Optimization, RWTH Aachen University, Aachen, Germany
[c] RWTH Aachen University, Chair for Fuel Cells, Faculty of Mechanical Engineering, 52062, Aachen, Germany

ARTICLE INFO

Dataset link: https://github.com/FZJ-IEK3-VSA/ActivityAssure, https://doi.org/10.5281/zenodo.10835208

Keywords:
Activity validation framework
Activity profile
Occupant behavior model
Load profile generator
Energy-related occupant behavior
Time use survey

ABSTRACT

The simulation of human behavior is an essential component in the domain of energy demand modeling. However, due to its diverse nature, it is often unclear whether a simulated behavior pattern is fitting to certain contexts. Combined with the poor availability of appropriate activity data, this makes proper validation of behavior models difficult. Existing validation approaches are limited and specialized to the respective use case, and are therefore not reusable or comparable. To address this issue, the new open-source framework ETHOS.ActivityAssure is presented for evaluation of generated activity profiles, supporting the validation of behavior models. For this purpose, an aggregated activity dataset is created from restricted European time use data and published to ensure reusability. Validation is conducted through a set of indicators and comparative plots, taking into account activity duration, frequency, and time. A categorization of person types and activities enables mapping to result categories commonly used in behavior models. The framework's capabilities are demonstrated on the residential demand model LoadProfileGenerator. The developed framework allows for consistent and reproducible validation of synthetic activity profiles targeting Europe, without requiring access to confidential data. This offers the opportunity to enhance both new and existing behavior models by identifying flaws, compare multiple modeling approaches, and thoroughly evaluate model quality for diverse target purposes.

1. Introduction

Many models across various domains incorporate human behavior in some way. Especially in transportation or energy research, the inclusion of accurate activity schedules is often essential, as appliance loads, building energy demands and traffic flows depend directly on human behavior [1]. None of these aspects are negligible, with appliance loads for example having a significant impact on residential energy demand. Thus, to build an accurate energy demand model, obtaining suitable activity schedules is essential. Accessible sources of this kind of data are, however, scarce. On the large scale, activity schedules are collected in time use surveys (TUSs). These are sets of activity diaries, usually for single days and from participants in different socio-economic situations. TUSs are conducted in many countries, but due to the high effort they only include a limited number of respondents, and are repeated at large time intervals, such as once per decade. Furthermore, as they contain detailed behavior data on the individual level, they are subject to privacy concerns and access is often restricted, even for anonymized datasets. This mostly leaves no choice but to fall back on behavior simulation [2].

In this study, the focus is on applications in the energy domain. This comprises two main types of models, namely occupant behavior models and load profile generators. Occupant behavior modeling is a growing domain that investigates the influence human behavior has on building energy performance, i.e., the energy demand required for usage of a building, including heating and cooling [3,4]. Load profile generators aim to produce load curves for various types of demand. In private households, a large share of energy consumption is caused by plug loads, human device usage is an integral part of these [5]. Naturally, there exists an overlap between these two model types, and many current models attempt to achieve a holistic representation of residential demand, including electricity, water consumption, heating, and other load types. For these applications, obtaining accurate activity schedules is crucial.

Approaches to acquire such schedules are manifold and depend on the research question. There are stochastic and deterministic models,

* Corresponding author at: Forschungszentrum Jülich GmbH, Institute of Climate and Energy Systems – Jülich Systems Analysis (ICE-2), 52425, Jülich, Germany.
E-mail address: d.neuroth@fz-juelich.de (D. Neuroth).

https://doi.org/10.1016/j.enbuild.2024.115036
Received 24 June 2024; Received in revised form 6 November 2024; Accepted 7 November 2024
Available online 13 November 2024

Table 1
Overview of common techniques for occupancy and activity profile validation in the literature, including the plots and numerical indicators used for comparing validation data and model data.

Author	Validation Data	Plots			Indicators		
		P1	P2	P3	I1	I2	I3
Richardson et al. [16]	UK TUS 2000	X					
Santiago et al. [17]	Spanish TUS 2010	X					
Aerts et al. [18]	Belgian TUS 2005	X				X	
McKenna et al. [19]	UK TUS 2000	X	X	X	X		
Nijhuis et al. [20]	Dutch TUS	X					
Wilke et al. [6]	French TUS 1998	X			X		
Kamara-Esteban et al. [21]	Smart Home measurements		X				X
Reynaud et al. [9]	French TUS 2010					X	
Bottaccioli et al. [7]	Italian TUS 2014	X			X		
Mueller et al. [22]	German TUS 2013	X	X				
Ziegler et al. [23]	German TUS 2013	X					
Kleinebrahm et al. [10]	German TUS before 2010, Mobility Data	X		X			
Chen et al. [8]	American TUS 2013–2017	X	X	X	X		
Osman et al. [24]	Canadian TUS 2015	X		X	X		

P1: Daily probability curves.
P2: Frequency distribution.
P3: Duration distribution.
I1: Indicators from daily probability curves.
I2: Clustering.
I3: Frequency and duration within confidence interval.

with different methods for generating diversity in behavior, respectively. Further notable features include model granularity, i.e., whether behavior is simulated at the individual or household level, and consideration of dynamic environmental influences. However, most behavior models rely on activity profiles from TUS data for calibration. This data is typically used to build Markov chains or probability distributions, which can then be used to simulate activity choices. Examples of this method include the models by Wilke et al. [6], Bottaccioli et al. [7], or Chen et al. [8]. SMACH by Reynaud et al. [9] instead consists of an agent-based model with preliminary activity schedules derived from TUS data, which in turn influence activity probabilities for the simulated agents. Kleinebrahm et al. [10] pursue yet another approach by training transformer models and long short-term memory-based neural networks on TUS and mobility data for a more realistic representation of weekly patterns. Other approaches do not rely on TUS data though. The LoadProfileGenerator, for example, a residential demand model by Pflugradt et al. [11], works with deterministic activity simulation, which is based on a set of continuously increasing desires that each person must fulfill through matching activities.

Due to the large amount of possibly relevant influences, the realistic and precise simulation of human behavior is a difficult task [12,13]. Hence, many behavior models are complex and elaborate. This emphasizes the necessity of thorough and all-encompassing validation. For this purpose, model developers can check created activity profiles or, where applicable, directly validate the final model output, such as load profiles. Due to the aforementioned difficulties regarding human behavior data, most authors focus on the latter component. Furthermore, many models are built upon a specific set of behavioral data, and so using the same dataset for validation is of limited use.

However, models increasingly aim to represent human behaviors more explicitly and in more detail. In order to support this approach to modeling, to evaluate result quality, and to enable further improvements, it is necessary to implement dedicated activity profile validation. This paper attempts to address this deficit by developing the generic validation framework ETHOS.ActivityAssure [14] that allows the adequacy of activity profiles to be verified using a large European behavior dataset [15]. With this framework, models can more easily be compared and evaluated, providing opportunities for reuse and improvement. ETHOS.ActivityAssure is part of ETHOS, the Energy Transformation Pathway Optimization Suite developed by the Institute of Climate and Energy Systems, Jülich Systems Analysis (ICE-2) at Forschungszentrum Jülich.[1] ETHOS provides models and tools for energy system analysis with a high level of detail.

The remainder of this paper is structured as follows: in Section 2, related human behavior models and the validation methods they employ are introduced. In Section 3, the ETHOS.ActivityAssure framework and the associated dataset are presented and explained. The application of the framework for the validation of a typical behavior model is described in Section 4. Finally, Section 5 concludes with a summary of the utility of the framework and possibilities for future improvements.

2. Related work

Up until now, no dedicated framework for the validation of activity profiles and corresponding models could be identified. Therefore, validation methodologies conducted by the authors of behavior models are reviewed instead. However, as only a few authors validate their generated activity profiles and instead focus on other model outputs, only a small number of methods for behavioral validation have been described thus far. Amongst these, there is a strong tendency towards visual comparisons of measured and generated data. Some, however, also employ numeric indicators. Table 1 presents an overview of common validation techniques in the literature.

2.1. Means of comparison

Comparisons of simulated and measured data are mostly carried out visually. By far the most common chart type used for this is daily probability curves. These curves are plotted for each activity, and indicate the probability of this activity being carried out at different times of the day. They enable the precise evaluation of time-dependent activity patterns and thorough comparisons of measured and generated data. This type of plot is predominantly used for occupancy models, which simulate the general state of occupants, including the presence and level of activity, but not the activities conducted [18–20]. Richardson et al. [16] and Santiago et al. [17] also employ slightly different variants of daily probability curves, demonstrating the possibility for state changes or for non-empty dwellings. In addition to the visualizations, some numerical

[1] ETHOS Model Suite: https://www.fz-juelich.de/en/ice/ice-2/expertise/model-services.

indicators were derived from the plotted curves. For that, common distance metrics including mean absolute error (MAE), root mean squared error (RMSE), and bias are used (Equations (1) to (3)). Unlike the MAE, the RMSE places more weight on large deviations. The bias has its own distinct meaning, as positive and negative errors cancel each other out in its calculation. When applied to daily probability curves, it represents the overall amount that an activity is carried out by too much or too little, neglecting any temporal shifts. The Wasserstein distance (Equation (4)) is another common metric, but one that has not yet been applied to this purpose, despite its unique properties. It represents the effort required to transform one curve into another. Hence, it can result in lower, more appropriate error values in cases of slight temporal shifts of large peaks, which usually lead to high MAE or RMSE values. In addition to these metrics, the Pearson correlation coefficient (Equation (5)) can be used as a measure of similarity and assesses the shape of the curves, ignoring the heights of peaks.

$$mean = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{1}$$

$$rmse = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

$$bias = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \tag{3}$$

$$wasserstein = \int_{-\infty}^{+\infty}|X - Y| \tag{4}$$

$$correlation = \sum_{i=1}^{n}\frac{(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{5}$$

x and y denote a daily probability curve of simulated and measured data, respectively. X and Y are the corresponding cumulative distribution functions.

Every individual plot and indicator alone only has limited validation capabilities, and so multiple different measures are used in combination. Most validations include at least one measure each to examine the average time, duration, and frequency of activities, as these are the basic characteristics of behavior profiles.

2.2. Level of aggregation

All measures depend on the level of aggregation, pertaining to whether a simulated population is validated as a whole or split into multiple categories to more accurately investigate differences in behavioral patterns. Most authors validate weekday and weekend profiles separately. Wilke et al. [6] distinguish between working and non-working people, whereas Ziegler et al. [23] combine both approaches. Compared to the average number of possible model configurations, this is, however, still a fairly rough categorization, as most models offer multiple options for age, sex, employment status or household composition of the simulated persons, which all affect activity profile generation. A second dimension of aggregation relates to the set of available activities. Each model defines a set of these, usually between five and 50, that make up the generated schedules. For the validation, however, some authors merge activities to groups to keep the effort manageable.

2.3. Examples of behavioral validation

In the following, some specific validation methods are presented in greater detail. Wilke et al. [6] show the daily probability curves of all activities together in a stacked line chart. In order to highlight deviations, the difference between TUS and the simulation results is also plotted in another line chart. As the authors use the same dataset for both the calibration and validation of the model, they apply ten-fold cross validation to still obtain meaningful results.

Kamara-Esteban et al. [21] calculate confidence intervals for expected values of activity frequencies and durations and determine whether the simulation results are reflected within. Reynaud et al. [9] propose a more generic methodology. The combined dataset of validation data and simulation results is clustered to see if the datasets can be easily distinguished. If not, this is seen as an indicator as to the validity of the results. In a follow-up, they add manual activity schedule verification by persons who have been recreated within the model, which is difficult to quantify and not scalable [25].

Bottaccioli et al. [7], Ziegler et al. [23], and Chen et al. [8] all make use of daily probability curves, but with different levels of aggregation. Ziegler et al. [23] created plots with 95% confidence interval curves to emphasize the extent to which results and validation data match. Chen et al. [8], on the other hand, complement their daily probability curves with histograms for activity count and duration. Other means of validation used by Mueller et al. [22] and Kleinebrahm et al. [10] include a histogram, which shows the overall number of activity changes per day, as well as plots showing the total time spent per activity. As Kleinebrahm et al. generate activity profiles with a length of more than one day, they also utilize auto-correlation and Hamming distance to assess diversity within a single profile. For single-day profiles, however, this approach is not applicable.

2.4. Evaluation of validation approaches

All of the presented validation methodologies are based on TUS or comparable data. In each case, only a single dataset was used, which was always the one the model had been calibrated with. This could be due to the additional effort that would be required to process different, not standardized, time use surveys, as well as due to the generally bad accessibility of these datasets. Given that, of the mentioned works, only Mueller et al. [22] and Ziegler et al. [23] used the same dataset, aside from McKenna et al. [19] and Richardson et al. [16], who worked on the same model, this circumstance poses an obstacle to model comparisons.

In summary, the methods applied for behavior model validation are highly heterogeneous with respect to validation data, comparative plots, indicators, and categorizations of person types and activities. The ETHOS.ActivityAssure framework introduced in the following section aims to combine the most relevant methods into a single, reusable and reproducible validation workflow with a more detailed categorization. In this way, it will help bring about precise, consistent, and meaningful validations, ultimately allowing better evaluations and comparisons of behavior models.

3. Methodology

One of the core issues the new validation framework is meant to solve is the poor accessibility of activity data. To that end, a dedicated validation dataset has been generated and published [15], and is presented in Section 3.1.

To aid in the validation process, a software framework was developed [14] that can be used to calculate the required statistics for input activity profiles and compare them to the validation dataset. This reduces the overall validation effort and provides a consistent set of visualizations and comparison indicators, as well as enables the tooling for categorizing data.

In Section 3.2, the creation of the validation framework is described in more detail. Section 3.3 details a selection of the indicators used for the comparison.

3.1. Creation of a validation dataset

In order to solve the data accessibility issues, the primary purpose of the generated dataset is to avoid disclosing individual information, so

Table 2
Person categorization criteria and possible values as contained in the HETUS dataset.

Categorization criteria	Possible values
Country[1]	AT, BE, DE, EE, EL, ES, FI, FR, HU, IT, LU, NL, NO, PL, RO, RS, UK
Sex	Female, male
Employment type	Full-time, part-time, retired, student, unemployed
Day type	Working day, rest day

[1] The two-character country codes from Eurostat are used to indicate the country.

that it can be made publicly available without constraints. As a basis for this, the Harmonized European Time Use Survey (HETUS) from 2010 was used [26]. This dataset, provided by Eurostat, combines the results from 17 national surveys. It includes activity diaries for single days at a 10-minute resolution, as well as personal and household characteristics for each respondent. Activities are specified using a three-layer coding system with a set of 108 possible ones. Due to the personal information it contains, this dataset is not publicly accessible. The structure of the HETUS dataset is shown in Appendix A. HETUS data are collected once per decade, with the most recent available iteration being that of 2010. The results from HETUS 2020 are still undergoing processing and are anticipated to become available in 2027.

An important factor for both activity and load models is household level dependencies, such as shared meals or exclusive use of appliances. Although activity profiles are primarily generated for individual occupants [6,18], some models take such influences into account [5,7,11,25]. Ideally, an activity profile validation would therefore need to evaluate entire households to check for correlated behavior patterns. However, for this to be possible, the validation data would need to include activity diaries for each resident of a household for the same day. This type of data is rarely available. In HETUS, the exact date of diary recording is obfuscated for privacy reasons, so there is no way to verify that diaries belonging to the same household were recorded on the same day. Thus, activity validation at the household level is not feasible with the available data, and all validation is done at the individual level. This is consistent with existing approaches to activity validation in the literature.

The underlying idea for improving accessibility is to aggregate the confidential HETUS dataset just enough to preempt privacy concerns and thus permit publication, but still retain sufficient information on different facets of human behavior. To achieve this, a suitable categorization of the activity profiles is necessary. Selecting an appropriate categorization scheme is crucial, as within a category no further differentiation will be possible. Fine-grained categorizations offer more precise distinctions, but the category sizes must not become too small either. Eurostat prescribes a minimum cell size of 20 for data that is derived from HETUS and intended for publication. Consequently, statistics for categories containing less than 20 entries cannot be published. Most behavior models make internal use of characteristics of the modeled person or household for a better adjusted activity simulation. These characteristics are good candidates for categorization, as they were deliberately chosen to reflect behavioral patterns and allow the selective validation of the respective person types of a model with matching behavior data. The specific attributes taken into account differ for each model. As a compromise, a set of four categorization criteria, namely country, sex, employment type, and day type, were established, as these cover the most important distinctions. Possible values for each of these are shown in Table 2. In regard to the categorization of day types, there exist a multitude of related approaches, which predominantly consider weekdays and one to three distinct categories of weekend days. As these approaches disregard weekend work, vacation, and holidays, a straightforward differentiation between working days and rest days was selected instead, determined by the total hours of employed work on a given day.

Combined, these criteria result in a total of 340 theoretically-possible categories, or 20 categories per country. As no data was available for retired and unemployed people from the Netherlands (NL), eight categories were excluded. Another 29 had less than 20 entries and had to be excluded. These include mostly working days for unemployed or retired people, which is to be expected, as well as some part-time categories, for which in general little data was available. In summary, 303 categories remain for use. An overview of the cell sizes is shown in Fig. 1. Due to data protection requirements, a cell size of 50 is also indicated for all smaller cells. When accounting for entries that are restricted from publication, as well as records that could not be processed due to missing information, 95.4% of the HETUS dataset could be incorporated into the developed dataset. The authors regard this as a sufficiently large share for realistically depicting relevant behavioral patterns. Appendix B presents the distribution of diaries utilized for the validation dataset by country.

As not all models make use of each of these criteria, it can be helpful to be able to merge arbitrary categories. For example, if a model only distinguishes between the working and non-working populations, separating retired and unemployed persons in the dataset might be impractical. However, for realistic results, the correct proportions should be considered when merging different categories. Otherwise, rarer activity profiles could have too much influence. For this purpose, the HETUS dataset provides an appropriate weight for each activity diary, taking both the person and date into account. The summed weights are stored for each category as part of the validation dataset, allowing dynamic merging for validating models without implementing all of the predefined criteria.

The 108 different activities of the HETUS data are unwieldy for comparison with common behavior models. Therefore, a single, general set of activity groups was devised, which combines all activities from HETUS, as well as from every reviewed model that specified its respective activities. To do that, similar activities were matched, and small and insignificant activities were iteratively merged into groups, until ultimately each activity group had an overall time share of $>2\%$ in at least one profile category. The final set of activity groups consisted of "clean," "cook," "dishwashing," "eat," "education," "iron," "laundry," "PC," "personal care," "radio/music," "sleep," "TV," "work," "not at home" and "other". A definition of each activity is provided in Appendix C. For all reviewed models in Section 2, including the LoadProfileGenerator, a mapping of their respective activity set to the merged set is provided as part of the framework.[2] This enables straightforward comparisons of the different models.

For each cell, all activity diaries are reduced to a set of statistics, each calculated per activity group. These include activity durations, frequencies per day, and activity probability profiles. Thus, the core aspects of behavior, activity frequencies, durations and times, are all taken into account. The validation dataset is freely available on Zenodo at https://doi.org/10.5281/zenodo.10835208.

3.2. Development of the validation framework

In order to support the validation of a behavior model using the new validation dataset, the ETHOS.ActivityAssure software framework was

[2] Location of the activity mappings within the framework repository: https://github.com/FZJ-IEK3-VSA/ActivityAssure/blob/main/activityassure/activities.

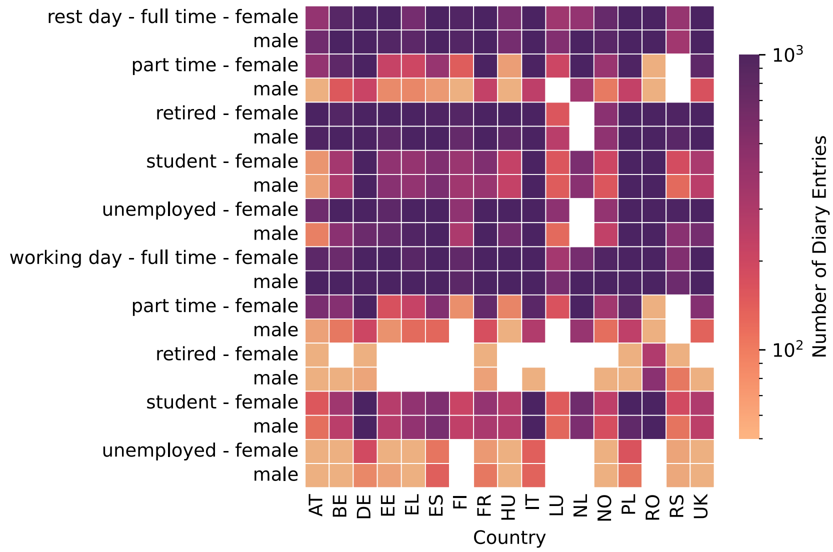


Fig. 1. Number of entries per category. Note that a size of 50 is specified for all cells with less than or equal to 50 but more than 20 entries, due to Eurostat requirements. White cells denote categories with less than 20 entries, which had to be excluded.)

developed [14]. It consists of two separate parts—a statistical calculation and a graphical validation user interface. Through this architecture processing of the model output data is decoupled from interactive validation.

As input, the activity profiles to be validated must be provided in the form of .csv files, each file containing a profile of arbitrary length for one person. In addition, a single .json file must contain each person's country, sex, and employment type. Finally, an activity mapping must define the translation of the activities used in the model to the activity set defined in Section 3.1, in another .json file.

In essence, the first part of the framework applies a similar transformation to this input data, as was utilized in the case of the HETUS dataset to obtain the aggregated validation dataset, combined with additional preprocessing. This encompasses translating activities in the profiles into the predefined groups according to the user's mapping, resampling to the 10-minute time resolution used in the HETUS data, splitting activity profiles spanning multiple days into 24-hour profiles and categorizing them. Ultimately, this component of the framework stores statistics per category of the processed data in the same file structure as the validation dataset. Instructions for this process, a sample script, and sample input files to show the exact file format are included in the repository [14].

Building upon that, the second component incorporates the validation and input data statistics and generates comparative figures and indicators in order to assess similarities. This portion was built as a web application using Dash [27] to allow for the interactive investigation of the results. Additionally, the framework can generate overview files, including all indicators for automatic processing. By default, the input data statistics are compared to validation statistics of the same profile category, meaning the same person and day type. In this way, it is possible to identify which person types are better represented by the model than others. However, for more exploratory purposes, comparisons of different categories are also possible. Comparison indicators and plots are shown as totals for the entire category, as well as for each activity group individually. This facilitates understanding of possible differences in behavioral patterns and the identification of the affected model components that might need to be adapted for better results.

In summary, the steps for using the framework include adding descriptive information to the input data, executing the data processing function to generate statistics, and then interactively validating the data with the web application.

3.3. Selection of comparison plots and indicators

The comparison of the validation and input data relies on comparative indicators and plots. For that, some of the different validation methods presented in related work were evaluated. The main emphasis was placed on daily probability curves, as these are widely used and provide information on the most relevant aspects of activity profiles. For a first, general overview, a stacked probability curve plot shows all activity groups at once, so as to be able to quickly identify the largest deviations. Additionally, individual probability curves for each activity group allow for a more detailed assessment to be made. In addition to that, histograms of activity repetitions per day and of activity durations are generated, as information on these aspects also characterizes human behavior but cannot be obtained from the daily probability curves.

For the calculation of comparison indicators, the common approach of calculating distance metrics and similarity indices from daily probability curves was applied. For that, the indicators MAE, RMSE, bias, Wasserstein distance, and Pearson correlation coefficient were selected, as defined in Section 2.1. As already noted, capturing the quality of behavior modeling in a simple indicator is difficult and usually visual comparisons using plots are required. Hence, the main purpose of these indicators is to quickly find the worst results of a behavior model, e.g., the person configuration or activity with the highest deviation, in order to review it manually.

In that regard, the four distance metrics of the common indicators (MAE, RMSE, bias, and Wasserstein distance) have a potential weak point. As they measure the deviation of two curves, their value depends on the value range of these curves. All probability curves and with them the metrics lie within the range $[0, 1]$, where 0 implies identical curves. However, the profiles for more time-intensive activities such as sleeping have generally higher values, and relatively smaller errors can result in larger metric values. This can lead to underestimations of errors for activities with a smaller overall time share, such as doing laundry. For instance, if laundry activities were entirely missing in the input data, the metric values would still be very small due to the low probability within the validation data. Whether or not that is desired depends on the application. Although some users might wish to obtain activity profiles that correctly incorporate the most important activities, others might require accurate profiles for specific activity types. To support both cases, all metrics were calculated in two variants, namely normal and scaled ones. For the scaled variant, the metrics were divided by the sum of the

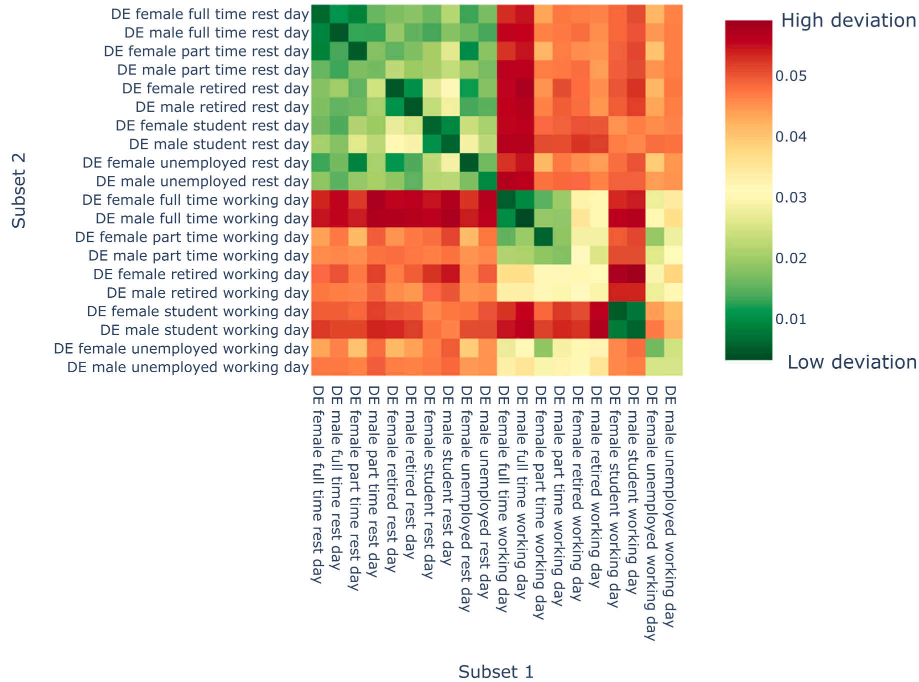


Fig. 2. Cross-validation of two HETUS subsets using the unscaled MAE, averaged across all activities.

probability curve of the validation data (see Equation (6)), leading to more comparable value ranges for different activity groups.

$$m_{scaled} = \frac{m_{normal}}{\sum_i y_i} \tag{6}$$

y denotes the daily probability curve of the validation data. The Pearson correlation coefficient was not affected by this.

For an initial test of the validation framework and the comparison indicators, a two-fold cross-validation was conducted. For that, the HETUS dataset for Germany was randomly split into two subsets, which were then treated as validation and input data and fed into the framework. The heat map in Fig. 2 shows the MAE values when comparing every possible pairing of activity profile categories. The darker main diagonal cells indicate that comparing the data of the same category results in better indicator values, as expected. Data from different categories instead produces worse values, especially when comparing working days to rest ones. This emphasizes the importance of differentiated validation for individual categories instead of a mere aggregated analysis.

4. Results

In order to demonstrate the usage and functionalities of the developed validation framework, the LoadProfileGenerator behavior model was applied [11]. The LoadProfileGenerator employs a psychological desire model to generate activity and load profiles for residential households. It is not calibrated with TUS data; instead, a number of typical German household compositions have been predefined using a small survey on typical activities and available appliances. For the validation, a set of activity profiles created with the LoadProfileGenerator was preprocessed, categorized, and compared to the validation data. Comparison indicators were calculated for each category, and were used to identify the model configurations that performed especially well or poorly. These were then examined in more detail using the web application in the framework.

The LoadProfileGenerator contains 157 persons across 66 predefined households, which cover most of the different household types in Germany. Each of these was simulated 50 times for a time span of one year, resulting in a total of around 2.6 million daily activity profiles. Of all 20 German activity profile categories, 18 are represented, as currently no male part-time employees are predefined, although configuring such persons is possible.

The validation was executed on a per-category basis. This means that, using the calculated behavioral statistics, all activity profiles from the LoadProfileGenerator within one profile category were compared to the respective statistics of the validation dataset. In case of deviations, single-person configurations of the LoadProfileGenerator were assessed individually by only using statistics calculated from the respective person. Section 4.1 briefly outlines how the validation indicators work in practice and compares category statistics to representative statistics for Germany. Next, after processing the profiles from the LoadProfileGenerator, some typical categories were selected and are depicted visually in Section 4.2.

4.1. Evaluation of validation indicators and categorization

As a first step, in order to check the practical usefulness of the selected validation indicators, heat maps were generated to show indicator values for different categories and activity groups. As can be seen in the example in Fig. 3, the MAE, RMSE, and Wasserstein distance behaved in the same way for the most part. Analyzing the exceptional cells for male and female students on rest days, it became apparent that, as expected, Wasserstein produced significantly lower error values when curves were shifted on the time axis. In this case, students in the LoadProfileGenerator data got up and went to bed too early. MAE and RMSE, however, were found to be very similar to each other.

In order to highlight the need for proper classification of person and day types, in contrast to a mere aggregated validation, behavioral statistics for a representative German population were created. Instead of individual statistics for each profile category, they contain just one set of statistics calculated from all German HETUS diaries, applying the respective diary weights included in the HETUS dataset. Comparing these average behavioral statistics to those of individual categories revealed significant differences. The most obvious deviation affects work, for which there is a clear differentiation between employees and unemployed or retired people. In addition to that, further activities are affected as well. For example, Fig. 4 shows how full-time employees mostly carry out "other" activities in the evening, after working hours, unlike unemployed people. Validating the activity profiles of any of

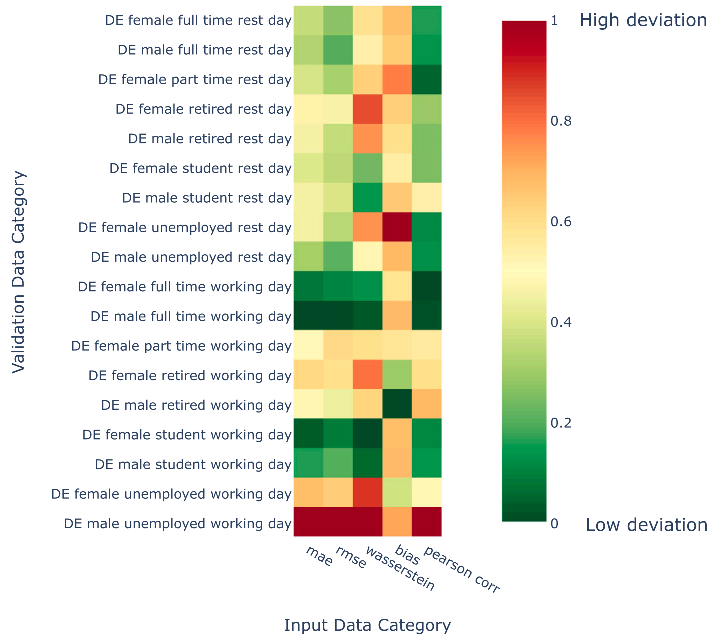
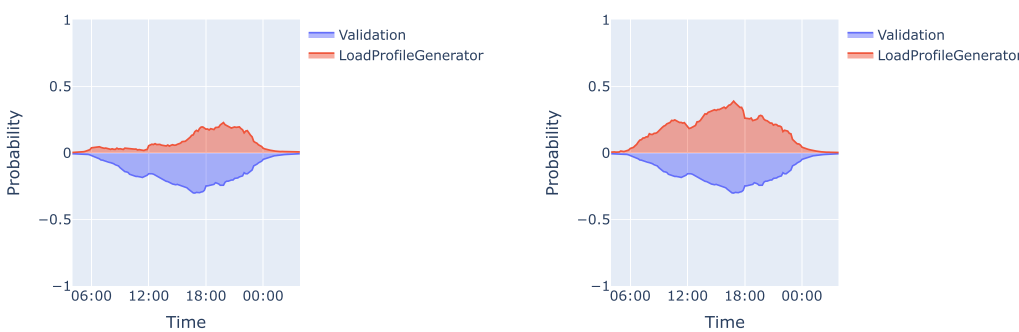


Fig. 3. Heat map of the five validation indicators, showing the averaged indicator values across all activities. Note that, for better comparability, all indicator values were individually normalized, and the value range of the Pearson correlation coefficient was reversed, so that lower values indicate a lower deviation. Normally, the Pearson correlation coefficient has a value range of $[-1, 1]$, with 1 and -1 indicating perfect positive or negative correlations, respectively, and 0 indicating no correlation. In this use case, however, the lowest observed values of the coefficient were just below 0. This was to be expected, as a negative correlation could only come about by coincidence.



(a) (b)

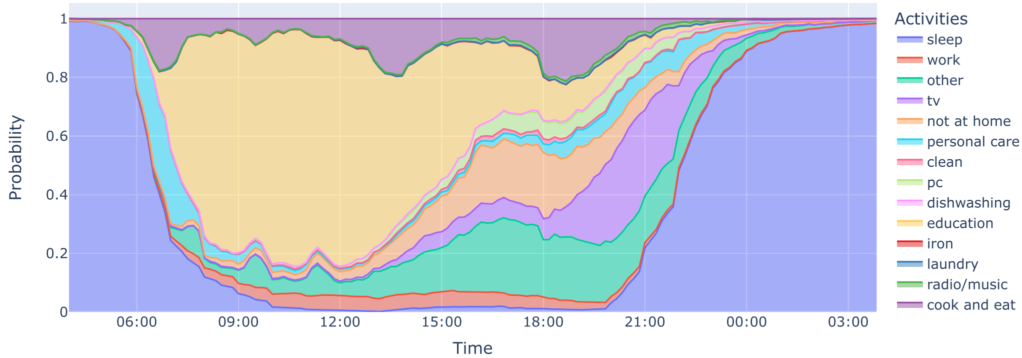
Fig. 4. Daily probability curves illustrating the likelihood of performing "other" activities at different times of the day for a female German full-time employee on a working day (a), and for a female German unemployed person on a rest day (b). Both curves are compared to the representative average for Germany (shown in blue). The signs of the probability values serve solely to differentiate the datasets and carry no additional meaning.

these types with average statistics is therefore not a sensible approach. Instead, individual statistics for each category are required to obtain useful results.

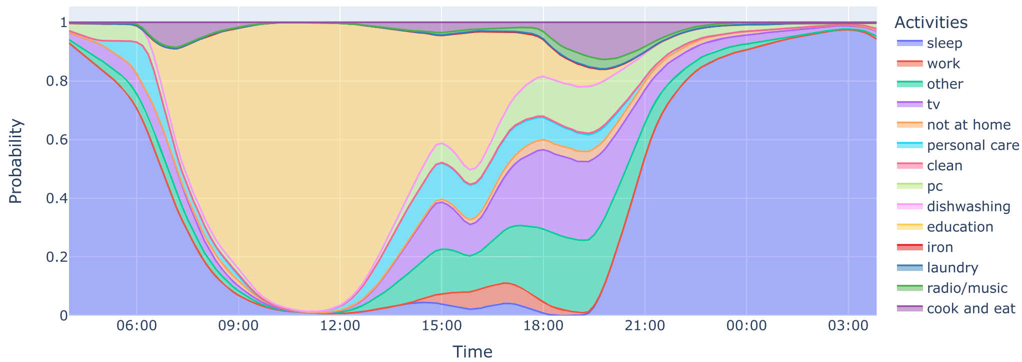
4.2. Assessment of individual categories

In the following, the utility of the ETHOS.ActivityAssure framework for validating behavior models is outlined for some typical categories. According to MAE, RMSE, and Wasserstein distance, one of the best-performing activity profile categories is that of female students on working days. This category is now analyzed in greater detail in order to provide an overview of the framework's capabilities. Fig. 5 shows the corresponding stacked probability curves of the validation data and of the LoadProfileGenerator. The low indicator values are primarily based on a good representation of basic activities such as sleep, as can be seen in Fig. 6. This is evidenced by the scaled variants of the indicators, which, within the category of female students on working days, are the lowest for the activities of sleep, education, and other. As unscaled indicator variants measure absolute deviations instead, they exhibit lower values for activities such as doing laundry and using a PC with less total time.

When looking at the histograms of the educational activity durations shown in Fig. 7, it becomes apparent that there are significant differences between the validation and model results. Although there are multiple flat peaks around the duration of 1:30h in the validation data, there are two clearly distinct ones in the data from the LoadProfileGenerator—at 1:00h and 6:00h. This discrepancy is in stark contrast to the good temporal distribution of educational activities, which closely matches the validation data, which suggests that there are frequent interruptions to educational activities in the validation data that are not accounted for in the LoadProfileGenerator. This pattern—fewer but longer activities—also appears in other activity groups, indicating that while the LoadProfileGenerator generally produces realistic schedules, its current level of detail may limit its suitability for applications that require precise activity duration modeling.



(a) Validation dataset



(b) LoadProfileGenerator

Fig. 5. Stacked daily probability curves for a female German student on a working day, for comparison.

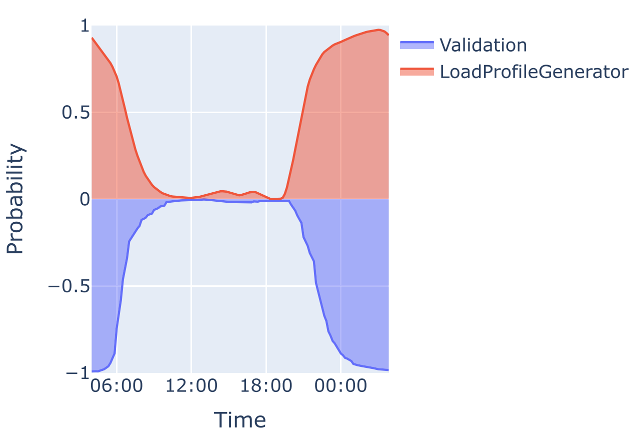


Fig. 6. Daily probability curves of the sleep of female German students on a working day.

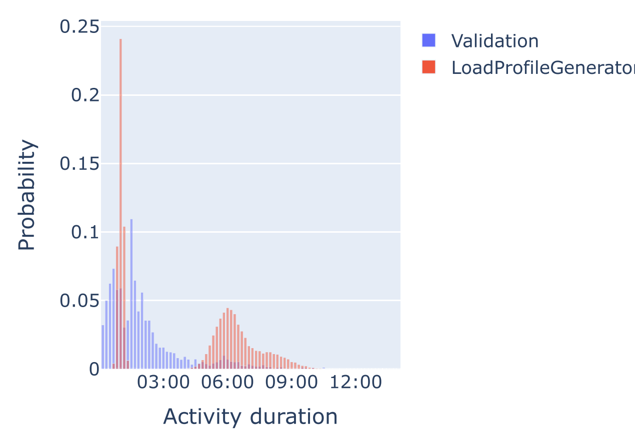


Fig. 7. Durations of the educational activities of female German students on a working day.

Overall, the day–night rhythm in the results of the LoadProfileGenerator is slightly shifted, which is also visible in Fig. 5. Some activities, including sleep, watching TV, and personal care, are well simulated, but the main peaks occur earlier than in the validation data. This is supported by a high Pearson correlation coefficient, e.g., 0.96 for sleep. Nevertheless, the total duration spent for the different activity groups is close to the validation data, with only one minute too long per day spent on personal care, for example. Time not at home is underestimated in the LoadProfileGenerator in terms of both duration and frequency, especially between 4:00 and 6:00 pm, indicating a lack of hobbies. The largest deviation was found in the early morning. Here, there were small peaks for several media usage activities within the LoadProfileGenerator that could not be observed in the HETUS data. In the latter, students spent most of their mornings focused on breakfast and personal care. Ironing was entirely missing in the LoadProfileGenerator activity profiles for female students, but as it has a very low share in the HETUS data with a maximum probability of less than 0.2%, this was negligible.

Another especially relevant group are full-time employees due to the high share of the German population they make up. They are, in general, well covered by the LoadProfileGenerator, in particular on working days. With respect to the category sizes, it became apparent that for some individuals, the share of working days was too small, with less than 150 per year. For a full-time German employee, a realistic number of

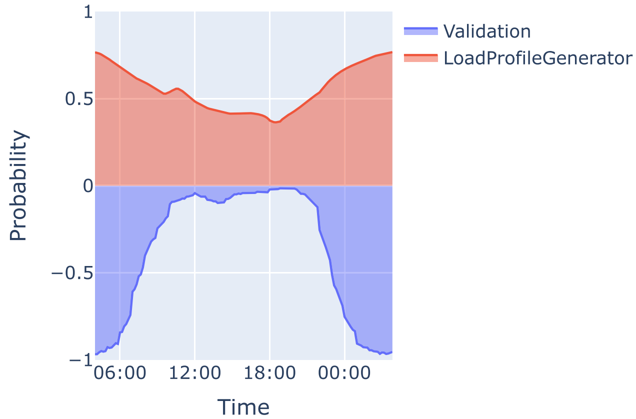


Fig. 8. Daily probability curves for the sleep of a male German unemployed person on a rest day, showing a high discrepancy due to an overrepresentation of irregular sleeping patterns in the LoadProfileGenerator.

working days would be around 200–230 per year. This large difference suggests an error in the affected person configurations of the LoadProfileGenerator. The activity profiles of all of the remaining individuals show realistic workday distributions.

This deviation does not affect the quality of the results for each day type. Using the scaled indicators, the activities that are best covered can be identified. These primarily include the important activities of work and sleep, which are both well-represented. For work, a similar phenomenon occurs as with the education activities of students, due to the missing integration of a lunch break. For the target use case of residential behavior modeling, this is negligible. The share of people watching TV is partially underestimated in the evening hours, and in turn PC usage is overestimated for several categories. These activities are similar in nature and compete for the same time period, but are not equally popular. The observed discrepancy indicates an inappropriate balance of these two activities in the LoadProfileGenerator, leading to the latter being selected too frequently.

Amongst the categories with the highest deviation is that of unemployed people. The main reason for the higher error lies in a bad match of the sleeping profiles, as can be seen in Fig. 8. Upon closer inspection of the activity profiles from the LoadProfileGenerator, it was found that this discrepancy is caused by only one of the archetypes of unemployed people in the LoadProfileGenerator, which is characterized by a highly irregular sleeping pattern. This pattern occurs in reality, but far less often than regular sleeping patterns, which are closer to those of employees.

4.3. Discussion

The example validations presented in Section 4.2 exemplify the possible usages of the ETHOS.ActivityAssure framework and demonstrate its capabilities. The framework offers a simple and straightforward method for conducting thorough comparisons of activity datasets and evaluating behavior models at a detailed level. The indicators enable categories or model configurations that warrant further investigation to be quickly found, allowing to focus effort on the most relevant aspects. As a core contribution, the validation dataset provides public access to behavioral statistics that are normally restricted. This enables meaningful comparisons of different models, the testing of their applicability to other countries, and, perhaps most importantly, validation with data that the model has not been calibrated with. This constitutes a significant advancement over the previous validation methodologies described in the literature.

The application of the framework helped to detect inaccuracies in the LoadProfileGenerator and identify specific components with the potential for further improvement. In this manner, the framework can facilitate improvements in the quality and realism of behavior models, ultimately leading to more accurate activity profiles.

It became evident during this process that a detailed, multifaceted examination of activity profiles is essential for identifying shortcomings and inaccuracies. Only looking at one aspect of behavior alone is not sufficient. This confirms the importance of incorporating many different indicators and plots into the framework and permit very specific deviations in behavior patterns, such as breaks interrupting working times, to be detected, leading to a more accurate estimation of the suitability of behavior models for different application scenarios.

Furthermore, it was found that within subgroups behavior patterns diverge considerably from the overall population average. This indicates that in order to achieve a meaningful validation, it is necessary to conduct a separate examination of each subgroup.

The framework is effective and beneficial for the developers of behavior models; however, there are also some points of criticism. As of now, the framework relies on a fixed categorization of persons, including in terms of country, sex, and employment status. Other attributes, such as household composition, have not yet been taken into account. Including these could allow for a more accurate categorization and enable the validation of more specific behavior patterns, but would also increase the demand for time use data.

Another drawback of categorization lies in the occurrence of small categories with statistics calculated from only a few activity profiles. This can potentially lead to distorted results for such categories, as statistical errors might increase in a small sample. These categories exhibit high indicator values and more visibly jagged probability curves. Although this can clearly be an obstacle for comparing modeling results across different categories, it only affects a negligible minority of the total population.

The clustering of activity profiles, as implemented by Osman et al. [24], represents a potential alternative to categorization. This method could facilitate a more precise matching of profiles and a more detailed identification of behavioral patterns, but it could impede the validation of behavioral archetypes.

An alternative approach to the presented framework might consist of the utilization of the few publicly-available TUS datasets. Building a framework on such data would enable a direct comparison of activity profiles instead of using aggregated indicators, which is not possible with the current method. In addition, the TUS dataset could be split into training and test data, permitting the calibration of multiple models with the same training dataset and then comparing their performance with respect to the test dataset. While this would be an ideal scenario for comparison of different behavior modeling methods, it is less suitable for assessing fully built and calibrated behavior models. Furthermore, the lack of publicly available TUS data would restrict such validation to a few regions, also making analyses of the applicability of models to other contexts more difficult.

As it only requires activity profiles as inputs, the framework can be applied to any behavior model or simulation tool that generates such profiles. The dataset published with this framework is only based on HETUS 2010 and is therefore limited in its applicability. Data for countries outside of Europe are not included, and although the dataset is from the latest available HETUS iteration, it is already outdated to a certain extent. In the meantime, there have been notable changes in occupant behavior, particularly in response to the COVID-19 pandemic. These include shifts in social behavior, increases in screen time, and other developments, some of which persisted beyond the end of the pandemic [28,29]. To date, none of the referenced models has been calibrated with data from after 2020; nevertheless, this caveat should be considered for new behavior models.

The open structure of the framework offers a potential solution to these issues. It is both feasible and desirable that new datasets, once available, undergo similar preprocessing and that they are then provided as additional validation datasets, whether for other regions or for new HETUS iterations.

One shortcoming of the presented methodology that cannot be resolved in this way is the lack of validation of household dependencies. As outlined in Section 3.1, the currently available data does not permit the analysis of such dependencies on a large scale. Consequently, the framework focuses on individual occupants. Should suitable data for this purpose become available in the future, another validation method and appropriate indicators must be devised.

5. Conclusions and future work

In this study, the new validation framework ETHOS.ActivityAssure for the consistent evaluation of activity profiles was developed. As an integral component, it includes an aggregated European behavior dataset to take measured activity schedules into account and still enable open, unrestricted reuse, in spite of strict data protection regulations. Unlike previous behavior model validations, this framework works with a fine-grained categorization to allow thorough and meaningful comparisons, only with matching data. The framework was shown to enable a very detailed assessment of a typical behavior model, the LoadProfileGenerator, and provide precise information on deviations in behavior patterns.

The framework and associated dataset are freely accessible and can be applied to any behavior model to identify flaws and assess or improve the accuracy and suitability of the results. Therefore, the behavior modeling domain, as well as domains that rely on it, such as demand profile generation, building performance simulation, an mobility modeling, can benefit from this framework and its deeper insights into the realism of generated activity profiles.

The described method for aggregating a confidential dataset was devised specifically to meet the data protection requirements from Eurostat, but it is expected to be applicable to other time use surveys as well, given some adjustments if needs be. This opens up the opportunity of creating more validation datasets in the future, such as from the Multinational Time Use Study [30], which contains time use surveys from different countries and years. In that way, the possibilities for behavioral validation can be extended, potentially covering more countries and an overall larger amount of data.

Disclaimer

This work is based on data from the Eurostat, Harmonized European time use survey, 2010. The responsibility for all conclusions drawn from it lies entirely with the authors.

CRediT authorship contribution statement

David Neuroth: Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Noah Pflugradt:** Writing – review & editing, Supervision, Conceptualization. **Jann Michael Weinand:** Writing – review & editing, Funding acquisition. **Christina Büsing:** Writing – review & editing, Supervision. **Detlef Stolten:** Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Helmholtz Association as part of the program "Energy System Design." Furthermore, it received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 891943.

Appendix A. HETUS data structure (Table A.3)

Table A.3
Synthetic data sample to show the HETUS structure. Included are the record IDs, socioeconomic characteristics, and the activity codes for all 144 10-minute intervals of a 24-hour diary (MACT1–MACT144).

HETUS Attribute	Profile 1	Profile 2	Profile 3
Country	DE	FR	FR
Sex	0	1	0
Household ID	1	2	2
Person ID	1	1	2
Diary	1	1	1
Work Status	1	3	4
(More socioeconomic attributes)	...	...	...
MACT1	11	11	11
MACT2	11	11	21
...	...	...	...
MACT144	11	39	11

Appendix B. Country proportions in HETUS (Table B.4)

Table B.4
Rounded proportion of HETUS diaries from each country used for the validation dataset.

Country	Eurostat Country Code	Proportion [%]
Austria	AT	2.3
Belgium	BE	3.1
Germany	DE	8.0
Estonia	EE	2.9
Greece	EL	4.2
Spain	ES	5.7
Finland	FI	2.1
France	FR	7.7
Hungary	HU	2.5
Italy	IT	11.3
Luxembourg	LU	1.1
Netherlands	NL	2.9
Norway	NO	2.1
Poland	PL	22.1
Romania	RO	16.7
Serbia	RS	1.7
United Kingdom	UK	3.9

Appendix C. Activity definitions (Table C.5)

Table C.5
Definition of the 15 activity groups in the merged activity set, as used in the framework and the validation dataset.

Activity Group	Definition
clean	Cleaning the dwelling or garden
cook	Preparing food or drinks
dishwashing	Cleaning dishes, manually or with a dishwasher
eat	Eating, consuming food or drinks
education	Studying, homework, lectures, helping children with studying
iron	Ironing
laundry	Doing laundry, including washing and drying
not at home	Any not work-related activity taking place away from home, such as shopping, going to the cinema or theater, or doing sports outdoors
pc	Any computer related activities, such as programming, communicating online, or playing computer games
personal care	Washing, getting dressed, and other personal hygiene and self-care related activities
radio/music	Listening to the radio, to music, and other forms of audio consumption
sleep	Sleeping, napping, or staying sick in bed
tv	Watching television
work	Working, including related activities such as commuting and taking lunch break
other	All other activities not matching the previous activity groups, for instance, reading, accompanying children, receiving guests, or pursuing hobbies

Data availability

The ETHOS.ActivityAssure framework and the corresponding validation dataset are both openly available under permissive licenses and are free to reuse. The framework is available on GitHub at https://github.com/FZJ-IEK3-VSA/ActivityAssure, and the dataset is available on Zenodo at https://doi.org/10.5281/zenodo.10835208.

References

[1] F. Johari, G. Peronato, P. Sadeghian, X. Zhao, J. Widén, Urban building energy modeling: state of the art and future prospects, Renew. Sustain. Energy Rev. 128 (2020), https://doi.org/10.1016/j.rser.2020.109902.

[2] E. Proedrou, A comprehensive review of residential electricity load profile models, IEEE Access 9 (2021) 12114–12133, https://doi.org/10.1109/ACCESS.2021.3050074.

[3] Y. Zhang, X. Bai, F.P. Mills, J.C.V. Pezzey, Rethinking the role of occupant behavior in building energy performance: a review, Energy Build. 172 (2018) 279–294, https://doi.org/10.1016/j.enbuild.2018.05.017, https://www.sciencedirect.com/science/article/pii/S0378778818307576.

[4] G. Happle, J.A. Fonseca, A. Schlueter, A review on occupant behavior in urban building energy models, Energy Build. 174 (2018) 276–292, https://doi.org/10.1016/j.enbuild.2018.06.030, https://www.sciencedirect.com/science/article/pii/S0378778817333583.

[5] E. McKenna, M. Thomson, High-resolution stochastic integrated thermal–electrical domestic demand model, Appl. Energy 165 (2016) 445–461, https://doi.org/10.1016/j.apenergy.2015.12.089, https://www.sciencedirect.com/science/article/pii/S0306261915016621.

[6] U. Wilke, F. Haldi, J.-L. Scartezzini, D. Robinson, A bottom-up stochastic model to predict building occupants' time-dependent activities, Build. Environ. 60 (2013) 254–264, https://doi.org/10.1016/j.buildenv.2012.10.021, https://www.sciencedirect.com/science/article/pii/S0360132312002867.

[7] L. Bottaccioli, S. Di Cataldo, A. Acquaviva, E. Patti, Realistic multi-scale modeling of household electricity behaviors, IEEE Access 7 (2019) 2467–2489, https://doi.org/10.1109/ACCESS.2018.2886201.

[8] J. Chen, R. Adhikari, E. Wilson, J. Robertson, A. Fontanini, B. Polly, O. Olawale, Stochastic simulation of occupant-driven energy use in a bottom-up residential building stock model, Appl. Energy 325 (2022) 119890, https://doi.org/10.1016/j.apenergy.2022.119890, https://www.sciencedirect.com/science/article/pii/S0306261922011540.

[9] Q. Reynaud, Y. Haradji, F. Sempé, N. Sabouret, Using time use surveys in multi agent based simulations of human activity, in: ICAART (1), 2017, pp. 67–77.

[10] M. Kleinebrahm, J. Torriti, R. McKenna, A. Ardone, W. Fichtner, Using neural networks to model long-term dependencies in occupancy behavior, Energy Build. 240 (2021) 110879, https://doi.org/10.1016/j.enbuild.2021.110879, https://www.sciencedirect.com/science/article/pii/S0378778821001638.

[11] N. Pflugradt, P. Stenzel, L. Kotzur, D. Stolten, LoadProfileGenerator: an agent-based behavior simulation for generating residential load profiles, J. Open Sour. Softw. 7 (2022) 3574, https://doi.org/10.21105/joss.03574, https://joss.theoj.org/papers/10.21105/joss.03574.

[12] J. Malik, E. Azar, A. Mahdavi, T. Hong, A level-of-details framework for representing occupant behavior in agent-based models, Autom. Constr. 139 (2022) 104290, https://doi.org/10.1016/j.autcon.2022.104290, https://linkinghub.elsevier.com/retrieve/pii/S0926580522001637.

[13] L.G. Swan, V.I. Ugursal, Modeling of end-use energy consumption in the residential sector: a review of modeling techniques, Renew. Sustain. Energy Rev. 13 (2009) 1819–1835, https://doi.org/10.1016/j.rser.2008.09.033, https://www.sciencedirect.com/science/article/pii/S1364032108001949.

[14] D. Neuroth, N. Pflugradt, J.M. Weinand, D. Stolten, ETHOS.ActivityAssure, Forschungszentrum Jülich, Jülich Systems Analysis (IEK-3), https://github.com/FZJ-IEK3-VSA/ActivityAssure, 2024.

[15] D. Neuroth, N. Pflugradt, J.M. Weinand, D. Stolten, ETHOS.ActivityAssure Data Set, 2024, https://zenodo.org/records/10835209.

[16] I. Richardson, M. Thomson, D. Infield, A high-resolution domestic building occupancy model for energy demand simulations, Energy Build. 40 (2008) 1560–1566, https://doi.org/10.1016/j.enbuild.2008.02.006, https://www.sciencedirect.com/science/article/pii/S0378778808000467.

[17] I. Santiago, M.A. Lopez-Rodriguez, D. Trillo-Montero, J. Torriti, A. Moreno-Munoz, Activities related with electricity consumption in the Spanish residential sector: variations between days of the week, autonomous communities and size of towns, Energy Build. 79 (2014) 84–97, https://doi.org/10.1016/j.enbuild.2014.04.055, https://www.sciencedirect.com/science/article/pii/S0378778814003818.

[18] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, F. Descamps, A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison, Build. Environ. 75 (2014) 67–78, https://doi.org/10.1016/j.buildenv.2014.01.021, https://www.sciencedirect.com/science/article/pii/S0360132314000304.

[19] E. McKenna, M. Krawczynski, M. Thomson, Four-state domestic building occupancy model for energy demand simulations, Energy Build. 96 (2015) 30–39, https://doi.org/10.1016/j.enbuild.2015.03.013, https://www.sciencedirect.com/science/article/pii/S0378778815002054.

[20] M. Nijhuis, M. Gibescu, J.F.G. Cobben, Bottom-up Markov chain Monte Carlo approach for scenario based residential load modelling with publicly available data, Energy Build. 112 (2016) 121–129, https://doi.org/10.1016/j.enbuild.2015.12.004, https://www.sciencedirect.com/science/article/pii/S0378778815304436.

[21] O. Kamara-Esteban, G. Azkune, A. Pijoan, C. Borges, A. Alonso-Vicario, D. López-de-Ipiña, MASSHA: an agent-based approach for human activity simulation in intelligent environments, Pervasive Mob. Comput. 40 (2017) 279–300, https://doi.org/10.1016/j.pmcj.2017.07.007.

[22] M. Mueller, F. Biedenbach, J. Reinhard, Development of an integrated simulation model for load and mobility profiles of private households, Energies 13 (2020) 3843, https://doi.org/10.3390/en13153843, https://www.mdpi.com/781146.

[23] F. Ziegler, S. Seim, P. Verwiebe, Joachim Müller-Kirchenbauer, A probabilistic modelling approach for residential load profiles, Zenodo (2020), https://doi.org/10.5281/ZENODO.3689339, https://zenodo.org/record/3689339.

[24] M. Osman, M. Ouf, E. Azar, B. Dong, Stochastic bottom-up load profile generator for Canadian households' electricity demand, Build. Environ. 241 (2023) 110490, https://doi.org/10.1016/j.buildenv.2023.110490, https://www.sciencedirect.com/science/article/pii/S0360132323005176.

[25] Q. Reynaud, M. Schumann, F. Sempé, Y. Haradji, N. Sabouret, Multi-agent simulation of human activity for residential electrical load and demand forecasting and its application to electric mobility, in: CIRED 2021 - the 26th International Conference and Exhibition on Electricity Distribution, vol. 2021, 2021, pp. 2808–2812.

[26] Eurostat, Harmonised European time use survey, https://ec.europa.eu/eurostat/web/microdata/harmonised-european-time-use-surveys, 2010.

[27] Dash, Plotly, https://dash.plotly.com/, 2024.

[28] M. Trott, R. Driscoll, E. Irlado, S. Pardhan, Changes and correlates of screen time in adults and children during the COVID-19 pandemic: a systematic review and meta-analysis, EClinicalMedicine 48 (2022) 101452, https://doi.org/10.1016/j.eclinm.2022.101452.

[29] L. Lucchini, S. Centellegher, L. Pappalardo, R. Gallotti, F. Privitera, B. Lepri, M. De Nadai, Living in a pandemic: changes in mobility routines, social activity and adherence to COVID-19 protective measures, Sci. Rep. 11 (2021) 24452, https://doi.org/10.1038/s41598-021-04139-1, https://www.nature.com/articles/s41598-021-04139-1.

[30] J. Gershuny, M. Vega-Rapun, J. Lamote, Multinational Time Use Study, https://timeuse.org/mtus, 2020.