Using neural networks to model long-term dependencies in occupancy behavior

Using neural networks to model long-term dependencies in occupancy behavior

Max Kleinebrahm[a], Jacopo Torriti[b], Russell McKenna[c,d], Armin Ardone[a], Wolf Fichtner[a]

[a] Chair of Energy Economics, Karlsruhe Institute for Technology, Hertzstraße 16, 76187 Karlsruhe, Germany

[b] School of the Built Environment, University of Reading, Whiteknights, PO Box 219, Reading RG6 6AY, United Kingdom

[c] DTU Management Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[d] School of Engineering, University of Aberdeen, Aberdeen AB24 3FX, United Kingdom

[a] Corresponding author: Max Kleinebrahm, max.kleinebrahm@kit.edu, +49 721 608 44691

Abstract

Models simulating household energy demand based on different occupant and household types and their behavioral patterns have received increasing attention over the last years due the need to better understand fundamental characteristics that shape the demand side. Most of the models described in the literature are based on Time Use Survey data and Markov chains. Due to the nature of the underlying data and the Markov property, it is not sufficiently possible to consider long-term dependencies over several days in occupant behavior. An accurate mapping of long-term dependencies in behavior is of increasing importance, e.g. for the determination of flexibility potentials of individual households urgently needed to compensate supply-side fluctuations of renewable based energy systems. The aim of this study is to bridge the gap between social practice theory, energy related activity modelling and novel machine learning approaches. The weaknesses of existing approaches are addressed by combining time use survey data with mobility data, which provide information about individual mobility behavior over periods of one week. In social practice theory, emphasis is placed on the sequencing and repetition of practices over time. This suggests that practices have a memory. Transformer models based on the attention mechanism and Long short-term memory (LSTM) based neural networks define the state of the art in the field of natural language processing (NLP) and are for the first time introduced in this paper for the generation of weekly activity profiles. In a first step an autoregressive model is presented, which generates synthetic weekly mobility schedules of individual occupants and thereby captures long-term dependencies in mobility behavior. In a second step, an imputation model enriches the weekly mobility schedules with detailed information about energy relevant *at home* activities. The weekly activity profiles build the basis for multiple use cases one of which is modelling consistent electricity, heat and mobility demand profiles of households. The approach developed provides the basis for making high-quality weekly activity data available to the general public without having to carry out complex application procedures.

Keywords: activity modelling, mobility behavior, neural networks, synthetic data

1. Introduction

In the course of the decarbonisation of domestic heat demand, it is expected that a large part of the heat will be generated by electricity (e.g. through heat pumps) (Paardekooper et al. 2018). In order to decarbonise the mobility sector, the aim is to increase the amount of electric vehicles in the European union from 1.3 million in 2020 to at least 33 million by 2030 (Transport & Environment 2020). Due to the expected developments, fundamental characteristics will change in the course of energy demand in the household sector. Furthermore, the introduction of stationary and mobile electricity storage systems as well as stationary heat storage systems enable the storage of energy over periods of single days and therefore open up flexibility potentials in the residential sector, which can support the integration of fluctuating renewable energies. To evaluate these flexibility potentials, fundamental relationships that shape household energy demand must be understood.

Occupant behavior has been identified as having a significant impact on household energy demand (Steemers and Yun 2009). Therefore, there has been an increasing research interest in the field of behavioral modelling over the last years with the aim to explain dynamics in residential energy demand based on energy related activities (Torriti 2014, 2017). A large number of studies focus on the modelling of activity sequences of single households or individuals with the objective to describe occupant behavior on an aggregated level for socio-demographic differentiated groups (Aerts et al. 2014; Flett and Kelly 2016; Richardson et al. 2008; Wilke 2013). Time use data (TUD) are used as a data basis, which provide information on the temporal course of occupant activities over single days and are available for various countries in the form of population representative samples (Eurostat 2000). Based on occupant behavior, different approaches were developed that connect occupant activities with electrical household appliances and thus generate synthetic electricity demand profiles (Yamaguchi et al. 2018). The aim of these studies is to gain a deeper understanding of household electricity demand in order to e.g. be able to evaluate device-specific efficiency measures, time-dependent electricity tariffs or load shift potentials.

However, TUD only provide information on activity patterns of individual days, therefore longer-term dependencies in mobility behavior and energy relevant *at home* activities that extend over several days are not captured in existing TUD based models. Figure 1 compares the autocorrelation of power consumption data generated on the basis of TUD with measured power consumption data. The autocorrelation in the generated data is underestimated. Especially, dependencies between subsequent days (48 lags) are not properly reproduced by the examined models.

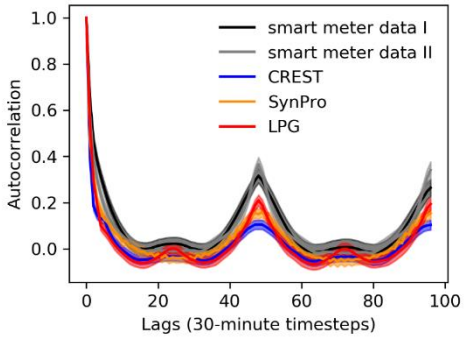


Figure 1: Mean autocorrelation and 95% confidence interval of electricity consumption profiles of the three load profile generators (LPG (Pflugradt 2016), CREST (Richardson et al. 2010), SynPro (Fischer et al. 2015)) and empirical smart meter data (I: HTW (Tjaden et al. 2015), II: (described in (Kaschub 2017))

Models based on device-specific power consumption data available over periods longer than one day are able to account for day-to-day variability in electricity demand (Yilmaz et al. 2017). However, due to the data underpinning these approaches, not much is known about the occupants and their

behavior, therefore it is not (easily) possible to calculate consistent heat and mobility demand profiles matching the electricity demand. One possible way to infer the occupancy behavior would be to use non-intrusive occupancy monitoring methods in order to calculate internal heat gains (metabolic gains and device-specific heat losses) (Chen et al. 2018). However, integrating demand through electrical vehicles would be another challenge.

The objective of this study is to develop a methodology that enables the generation of synthetic weekly activity schedules in which long-term dependencies in mobility behavior and energy relevant *at home* activities are captured on an individual level. These schedules can be used as a basis for generating consistent energy service demand profiles, taking into account heating, mobility and device specific energy service demand. In order to identify trends and potentials at the individual household level, like flexible charging behavior of electric vehicles, day-to-day variability in mobility patterns needs to be captured in the proposed approach. Therefore, novel machine learning based algorithms from the field of natural language processing (NLP) which are capable of capturing long-term dependencies in time series are transferred to the field of activity modelling. To answer the research question to what extent these algorithms are able to capture long-term dependencies in individual energy related occupancy patterns while maintaining the diversity of occupancy behavior on an individual and aggregated level, two behavioral data sets are combined in a two-step approach. Mobility data are used which provide information about weekly mobility patterns and combined with time use survey data which provide detailed information about daily activities (sleeping, cooking, eating, …).

The two-step approach enables to combine the advantages of mobility data (long-term dependencies in mobility behavior) with the advantages of TUD (detailed information about activities) and generates high quality weekly activity schedules. Novel machine learning algorithms which are used in the area of NLP are used for the first time to model occupancy behavior. These models have fundamental advantages over Markov chains, because they provide the capability to learn long term dependencies in time series. In comparison to existing approaches which were developed to reproduce aggregated occupancy behavior the proposed approach reproduces aggregated occupancy behavior and at the same time provides high quality individual activity schedules. Therefore, the synthetic activity schedules can be used to analyse trends in the household sector on an individual level and to examine their impact on an aggregated level at the same time. Due to the rich socio-demographic information in the underlying data sets, differences in behavior between socio-demographic groups can be analysed based on the synthetic activity schedules.

The paper is structured as follows. ***Section 2*** presents an overview about current approaches to activity based residential demand modelling and gives a short introduction to the field of social practice theory. Furthermore, the latest developments in the field of NLP are summarized. Section 3 presents the mobility and activity data used in this work. Subsequently, two autoregressive models are presented for the generation of weekly mobility schedules and two imputation models are presented for enriching the synthetic mobility schedules with energy related activity information. The section concludes with a presentation of the metrics used to evaluate the activity plans. In Section 4 the generated activity schedules are evaluated. Finally, the results are discussed and an outlook on future work is given in Section 5 before conclusions are drawn in Section 6.

2. Introducing NLP to activity modelling

The majority of studies in the residential energy demand modelling literature simulate residential energy demand based on activity patterns. The most important data basis for modelling activity sequences is TUD. TUD are large-scale surveys which provide detailed information about how people spend their time. The mean of data collection is the time-diary instrument in which the respondents enter their activities in regular time steps. These so-called time-diaries contain activity sequences for

the period of usually one single day. When selecting households for the study, care is taken to select a sample of households representative of the population. Time diaries are collected for all persons in the households except for young children for usually one weekday and one weekend day to capture the differences between the days. Since TUD are collected in a harmonised procedure in most countries in Europe, these data provide a good basis for a variety of similar models for modelling activity sequences. In the following, different model approaches are presented which generate activity sequences based on TUD and similar activity-based data sets. Furthermore, the weaknesses of the models reviewed in the literature is described and a short insight into social practice theory is given. Finally, the field of NLP is briefly introduced due to similarities in modelling human behavior and language.

2.1. Markov chain based approaches

One of the most commonly used approaches to map activity sequences is to describe them as Markov chains. A Markov chain is a stochastic process that describes a sequence of possible states in which the probability of each state depends only on the previous states. The state space of a Markov chain describes the set of possible states and their corresponding state transition probabilities. The abstract idea behind the modelling of activity sequences that describe the behavior of individuals is that individuals go about their lives by transitioning between different elements of a set of potential states of activity (Ramírez-Mendiola et al. 2019). Richardson et al. have developed an occupancy model which uses a first order Markov chain and distinguishes between the states 'active at home' and 'not active at home' for each person of a household (Richardson et al. 2008). Based on aggregated household states they calculate transition probabilities in order to model the activity level of the household over the timeframe of one day. By modelling households in an aggregated way instead of individual persons, inter personal relations are better represented than in models where individuals are modelled individually (McKenna et al. 2015). First order Markov models are adequately suited to describe processes that fullfill the Markov property. The term Markov property refers to the memorylessness of a stochastic process. For a first order Markov model, this means that the transition to a subsequent state depends only on the current state and is independent of previously observed states in the evolution of the process. It is obvious that residential activity schedules represent more complex processes and therefore cannot easily be represented by a first order Markov model. To overcome this problem, a variety of more complex Markov models have been presented in recent years. In contrast to first order Markov models, so-called semi-Markov models determine not only the subsequent state but also the duration of the subsequent state. As this kind of models represent an improvement to first order Markov Chains, due to a better mapping of state durations, they are used in various studies for activity modelling (Aerts et al. 2014; Wilke 2013; Bottaccioli et al. 2019). Flett et al. (Flett and Kelly 2016) present a Markov model for occupancy simulation that uses transition probabilities which are calculated based on the current state and the length of the current state. By considering the state length of the current state, this model represents an improvement over previous models, so that this model cannot be called memoryless. The logical next step would be to develop higher order Markov models, which allow any number of past states to be taken into account when choosing the subsequent state. However, two serious issues can be associated with higher-order Markov chains. On the one hand the number of free parameters in the model increases exponentially with the order of the model and on the other hand the collection of all possible full high-order Markov chain models is limited and completely stratified (Ramírez-Mendiola et al. 2019). Ramírez-Mendiola et al. (Ramírez-Mendiola et al. 2019) addressed this issues by presenting a Markov chain model with variable memory length which allows the order of the model to vary during the evolution of the stochastic process. In order to find relevant portions of the past based on the influence on the outcomes of the transition probabilities to subsequent states the authors present a novel algorithm based on the Kullback-Leibler divergence and the log-likelihood test.

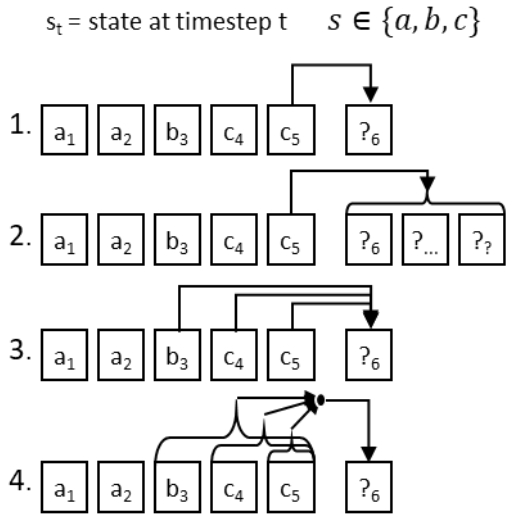


Figure 2: Graphical representation of the process of sequence generation with different kinds of Markov chains (1. first order Markov chain, 2. semi Markov chain, 3. higher order Markov chain, 4. Markov chain with variable memory length)

A graphical overview of the different Markov chain variations can be seen in Figure 2. It can be concluded that over the last few years more and more complex models based on Markov chains have been developed, which partly overcome the memorylessness problem. However, due to their structure, Markov models are only able to capture the states of the short-term past in order to predict subsequent states. Long-term relationships in daily schedules cannot be adequately represented by these types of models.

2.2. Timing of social practices

Markov chain approaches are based on the assumption that activities develop over time and are only dependent on the evolution of previous states. However, social practice theory literature points out that in order to understand people's daily/weekly schedules these should be treated as a whole (Shove et al. 2012; Torriti 2017). While practice theoretical accounts of social life vary, they remain consistent on at least two counts: (1) that practices are shared (socially/as part of the social i.e. performed by more than one person) and, because of that, (2) are repeated (performed more than once). If we also add that practices are connected and depend more and less on each other in being reproduced, it follows that we need to know more about how practices are repeated and with what effect for the relative strengths of their dependencies, connections, and extended relationships. In order to do justice to this statement in the patterning of activities, models must be developed which not only make it possible to capture connections between activities from the short-term past in order to predict the future, but also capture higher-level patterns which shape patterns of people's activities. In other words, models need to understand how temporal dynamics are embedded in the social world in order to understand how activities and thus energy consumption change and vary over time (Walker 2014). The majority of people structure their lives in daily rhythms, which are based on regular working hours, meal times and other constraints. These constraints form the basis for a certain degree of synchronization of social activities and thus for demand patterns (Walker 2014). Future models should be able to recognize and reproduce logical sequences in activity patterns, so that dependencies in activities are taken into account. For instance, food should first be prepared and then eaten.

Hilgert et al. (Hilgert et al. 2017) use a utility-based stepwise regression approach to generate weekly activity schedules for travel demand models. Due to the observation period of one week and the associated extended requirements for the mapping of activity sequences (day to day stability and variability of personal behavior), this approach differs from the approaches presented so far. Compared to Markov chain based approaches, activity sequences do not evolve over time but are the

result of regression based utility functions and time budgets. Based on Bowman (Bowman 1998) the construction process of activity schedules is split into smaller decisions due to the high complexity of constructing the entire schedule directly. These small decisions are then integrated downward vertically in the form of many logistic regression models. Due to the large number of regression models and their integration, many assumptions must be made when creating such a model, which increase the assumption bias. Future approaches should be less assumption driven to be easily transferable to different applications and datasets. To capture the high complexity of a complete activity plan without many intermediate steps, as described by Bowman, data-driven approaches could be used that need less assumptions and can capture complex relationships due to their structure. Table 1 gives an overview of the approaches presented in this section and compares them with the approach presented in this study.

Table 1: An overview of selected models for modelling occupancy behavior

Study	Database	Approach	Object of consideration	Country
(Richardson et al. 2008)	TUD	Markov - 1st order	Household	UK
(Wilke 2013)	TUD	Markov - semi	Individual	FR
(Bottaccioli et al. 2019)	TUD	Markov - semi	Individual	IT
(Aerts et al. 2014)	TUD	Markov - semi	Individual	BE
(Flett and Kelly 2016)	TUD	Markov - higher order	Individuals	UK
(Ramírez-Mendiola et al. 2019)	TUD	Markov - variable length	Individual	UK
(Hilgert et al. 2017)	MOP	Regression	Individual	DE
This study	**MOP + TUD**	**Neural networks**	**Individual**	**DE**

2.3. A brief review of natural language processing

The term natural language processing covers applications such as text classification, text understanding, text generation and text translation. NLP algorithms give machines the ability to read, understand and derive meaning from human languages. Over the last years NLP evolved from the era of punch cards and batch processing, in which the procession of a sentence could take up to 7 minutes, to the era of Transformer based model architectures like Googles BERT or OpenAIs GPT-3 with models up to 175 Billion parameters which are trained on large web corpora like Wikipedia and are able to generate articles which human evaluators have difficulty distinguishing from articles written by humans (Young et al. 2017; Brown et al. 2020; Devlin et al. 2018).

The first neural language model was based on a feed-forward neural network (Bengio et al. 2003). Vector representations of the n previous words are taken from a table and used as input in order to predict the probabilities of the following words. Nowadays dense vector representations of words or word embeddings are trained in an efficient way while training the neural network and are capable of capturing the context of words in a document (Mikolov et al. 2013).

From 2013 on neural network models in the form of recurrent neural networks (RNN), convolutional neural networks (CNN), and recursive neural networks got adopted in the field of NLP (Sutskever 2013; Kalchbrenner et al. 2014). RNNs are the obvious choice to deal with dynamic word sequences as they process the sequences from left-to-right or right-to-left and provide some kind of memory in the form of the hidden state (Elman 1990). RNNs in the form of long-short term memory networks (LSTM) proved to be more resilient to the vanishing gradient problem and therefore be able to better represent long-term dependencies in time series (Hochreiter and Schmidhuber 1997). The in 2014 presented sequence-to-sequence approach builds the basis for multiple machine translation applications. First, an LSTM-based encoder is used to compress an input sequence into a vector representation and then a decoder network, also based on LSTMs, predicts the target sequence step by step (Sutskever et al. 2014). The main shortcoming of the sequence-to-sequence approach is that

the input sequence needs to be compressed into a fixed-size vector. The Attention mechanism tackles this shortcoming by allowing the decoder to look back at the input sequence hidden states, which are provided as additional input to the decoder (Bahdanau et al. 2015). A rare feature of the Attention mechanism is, that it provides superficial insides about the learning process by providing information, through the attention weights, about which parts of the input are relevant for particular parts of the output. In 2016 Google presented their neural machine translation system which consisted of a deep LSTM network combining multiple encoder and decoder layers using residual connections and the attention mechanism (Wu et al. 2016). However, in 2017 the paper "Attention is all you need" was presented, which builds the basis for numerous transformer architectures which work on the principle of self-attention and define the state of the art in multiple NLP tasks (Vaswani et al. 2017; Brown et al. 2020). It was shown that the sequential nature can be captured by only using attention mechanisms and positional encodings without the use of RNNs. Due to the fundamental constraint of sequential computation of RNNs, it is not possible to parallelize training, therefore it is hard to learn on long sequences. Transformer models are fully based on fully connected layers and can be easily parallelized. Since 2017 multiple different transformer based architectures were introduced, consisting of multiple encoder and/or decoder blocks and an increasing number of trainable parameters (Wolf et al. 2020). In figure 3 the model architecture of a sequence to sequence RNN based model is compared to the model structure of an attention based transformer, consisting of an encoder and decoder block.

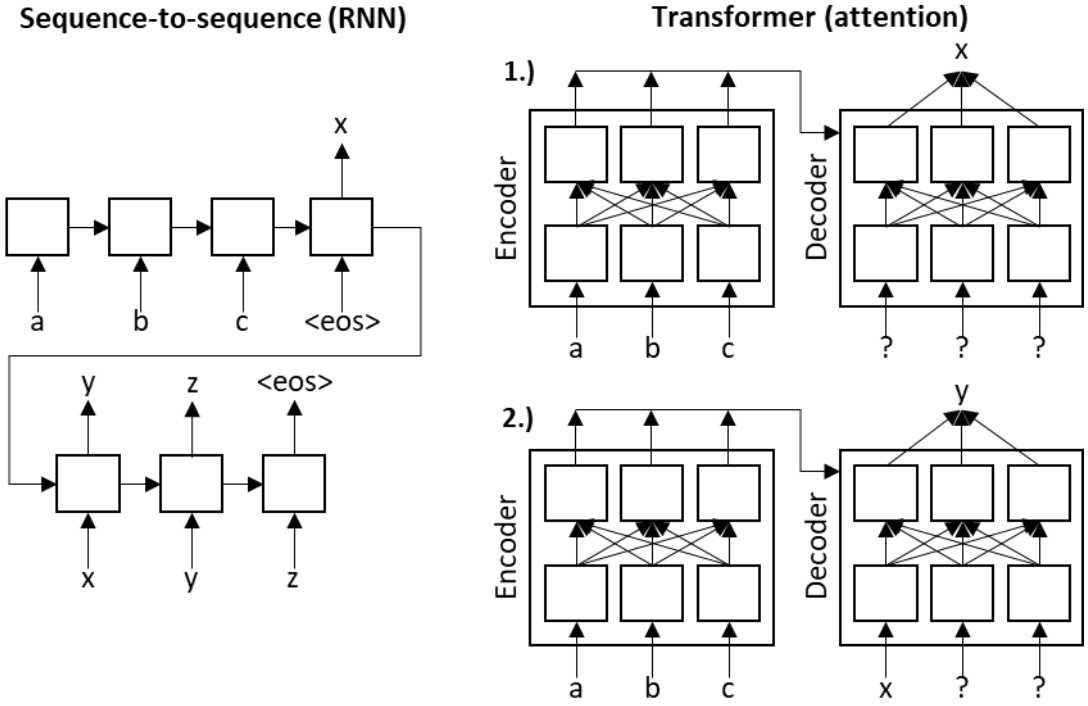


Figure 3: Abstract graphical representation of the RNN based sequence-to-sequence architecture (left) (Sutskever et al. 2014) and an encoder/decoder based transformer architecture on the right (Vaswani et al. 2017)

Adversarial learning methods have gained increased intention especially in the area of image processing/generation and have also been used in different forms in NLP over the last years. Generative adversarial networks (GANs) for example are able to generate synthetic data with similar statistical properties as real data by using two neural networks, a generator and a discriminator (Goodfellow et al. 2014). The generator produces synthetic data and the discriminator classifies generated data as fake and real data as real. Both networks are trained in an iterative way while trying to minimizes the reverse Kullback-Leibler divergence. Therefore, in comparison to the previously presented model architectures, GANs are not trained by maximum likelihood estimation (MLE) and thus are said to be less vulnerable to suffer from the exposure bias in the inference stage: the model generates a sequence iteratively and predicts next token conditioned on its previously predicted ones

that may be never seen in the training data (Yu et al. 2017). With that in mind many GAN based architectures were developed for natural language generation based on the approach presented in (Yu et al. 2017) which combines GANs with a reinforcement learning policy in order to deal with the differentiability problem. However, it was shown that MLE based approaches still dominate GANs when quality and diversity metrics are taken into account (Caccia et al. 2020). Therefore, GAN architectures are not considered further in this work, even if they form a promising basis for future work.

3. Data and Methodology

The German Mobility Panel (MOP) and German Time Use Data are used as an exemplary data source for analysing activity patterns in this study. In Section 3.1 the data preparation of the two data sets is described and the processed data is visualized. Further on, Section 3.2 presents the methodology developed to generate weekly activity schedules. Finally, Section 3.3 describes the metrics that are used to evaluate the activity plans.

3.1. Data

3.1.1. German mobility panel

The MOP collects information on the mobility behavior of the German population every year since 1994. About 1,500 to 3,100 persons (10 years and older), who make up about 900 to 1,900 households, fill out travel diaries over a period of one week. The travel diaries contain information about all trips during the week (start and arrival time, distance, modes used, purpose). In addition, socio-demographic information and information on refuelling behavior are recorded in the form of personal, household and fuel diaries. The survey is conducted every year in autumn to avoid distortions caused by holidays. The data is representative of the travel behavior of the German population. The Institute for Transport Studies at the Karlsruhe Institute of Technology is responsible for the implementation and design of the survey (Weiß et al. 2016; Zumkeller, Chlond 2009). Due to changes in the survey design, data from the surveys from 2001 to 2017 are used in this study.

3.1.2. German time use survey

For the analysis of energy relevant activities, the German part of the Harmonized European Time Use Survey, supplied by the German Statistic Office, was used (Destatis 2006; Eurostat 2000). Since the current version of 12/13 incorrectly recorded the location of the people, this data is not used. The data set contains activity diaries and socio-demographic information for 11,921 individual persons (age > 10 years) out of 5,443 households. Most of the participants provided diaries on two weekdays and one weekend-day in a 10-minute resolution. In this study time dependent data about primary activities and location as well as socio-demographic information are used.

3.1.3. Data preparation

In general, neural networks based machine learning methods have good adaptive feature learning ability. But in the present study the employed datasets are of a very different format, therefore they need to be aligned before the training. In order to create activity plans from the travel diaries, the basic dataset consisting of 833,986 travel entries for 35,014 person-weeks is converted into weekly activity plans with a time resolution of 10 minutes. The generation of activity plans is inspired by Hilgert et al. (Hilgert et al. 2017). In a first step, person weeks with missing or unrealistic entries are eliminated so that finally 26,610 person-weeks can be used for further analysis. Based on the travel entries and their purpose, states are determined for each time interval of the week. The choice of the initial state is based on the final state of the time series. Subsequently, the data are aggregated from a 1-minute

resolution to 10-minute resolution, assuming the state that is most frequently taken in the respective 10-minute interval. The reason for the reduction of the temporal resolution of the data is, on the one hand, the increased information density, since machine learning algorithms have problems with sparse data. On the other hand, TUD data are also recorded in 10-minute resolution.

The diary entries in the German TUD consist of more than 200 activity codes describing activities in the everyday life of human beings. Before the diary data is used as input for further processing, these activities are aggregated to activities relevant for household energy demand. The choice of activities is based on similar studies (Fischer et al. 2015; Richardson et al. 2010). The aggregated activities are visualized in Figure 4. In the upper two figures, the time course of the aggregated state probabilities of the two data sets is provided over a week. The lower two partial figures show example artificial activity plans for individual persons. Interday dependencies in behavior from Monday to Friday can be easily recognized from the visualization of the mobility schedule. The example activity plan, on the other hand, provides detailed daily information on energy-related home, sleep and mobility activities. Further comparative analyses based on socio-demographic characteristics of the data sets can be found in Section 5 and in the appendix.

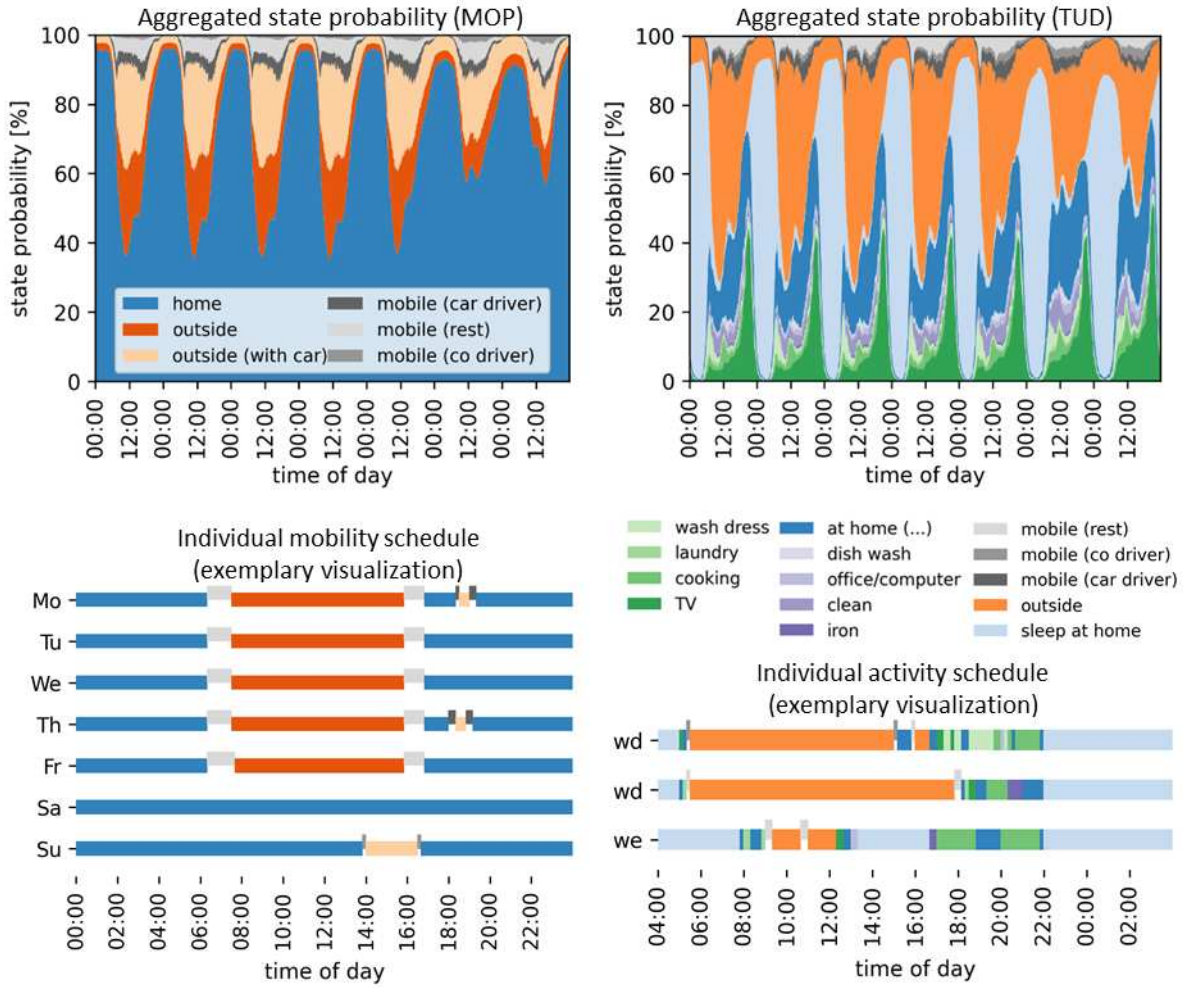


Figure 4: Visualization of aggregated state probabilities and exemplary artificial individual diary entries based on the MOP (Weiß et al. 2016) and the TUD (Destatis 2006)

3.2. Methodology

The approach for the generation of weekly activity schedules with a time resolution of 10 minutes is presented in Figure 5. In the first step, weekly mobility schedules of individual persons from the German Mobility Panel are used as input data. The objective of the first step is to generate synthetic

mobility schedules with statistical properties similar to the empirical schedules. The developed approaches are autoregressive. This means that it is assumed that the choice of the next mobility state ms_{t+1} only depends on all the states $ms_{0\ldots t}$ that have already been observed. In Section 3.2.1, an LSTM-based and an attention-based approach for sequence generation of mobility states are presented. ***Due to the similarity of the underlying problem, the selection of the methods used in this paper is based on the models that define the state of the art in the field of NLP. These are currently attention-based transformer architectures. Before that, LSTM based neural networks were used as described in Section 2.3.***

The objective of the second model step is to enrich the synthetic mobility plans with energy-related *at home* activities. For this purpose, two imputation models are presented in Section 3.2.2. Bidirectional LSTM model architectures are compared with attention-based architectures. Time Use Survey data from individuals are used to train the models. During the prediction process, the synthetically generated weekly mobility schedules are fed into the imputation model as input and the *at home* state is replaced by energy-relevant activities. A graphic representation of the step by step procedure of the autoregressive and imputation models can be found in Figure 7 a.

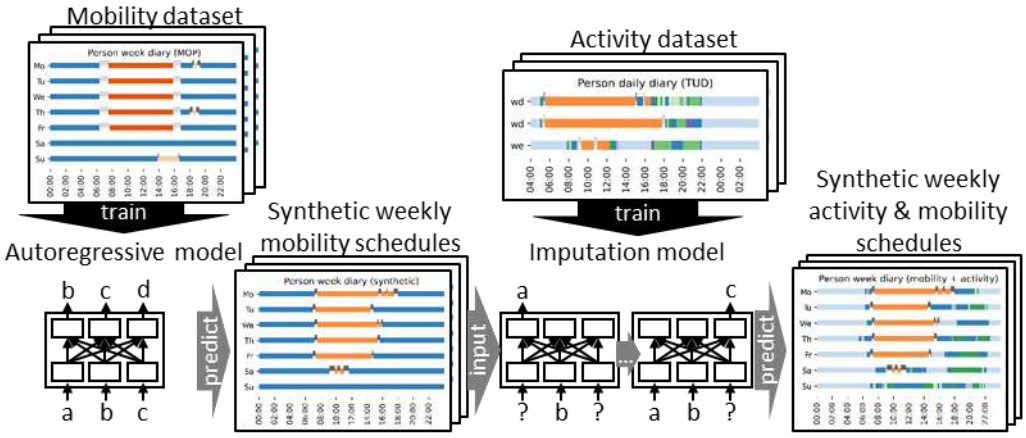


Figure 5: Two-step model approach for generating weekly activity schedules

3.2.1. Autoregressive models for weekly mobility schedule generation

To generate high-quality mobility plans on an individual level and at the same time representative mobility plans on an aggregated level that adequately describe the diversity of human behavior, approaches are required that capture the complex relationships in human behavior. In contrast to the Markov-based approaches used in the majority of the studies described in Section 2.1, LSTM and attention-based approaches can take into account longer-term time dependencies in the timing of individual activities due to their different memorisation mechanisms. While in Markov models probabilities are assigned to individual activity sequences and thus the number of free parameters increases exponentially with the order of the model, these kind of models are not suitable to take into account long-term dependencies in behavior between single days (Ramírez-Mendiola et al. 2019).

LSTM based models process time series sequentially and take as input the current state vector $x_t \in \mathbb{R}^d$ the hidden state vector $h_{t-1} \in \mathbb{R}^h$ and the cell state vector $c_{t-1} \in \mathbb{R}^h$. The dimension of the hidden state and the cell state vector h is the number of LSTM units which define the memory capacity of the LSTM cell. The cell states are adjusted every timestep using different gating mechanisms (input gate, output gate, forget gate) and activation functions. Due to the additive structure of the LSTM cells they partly solve the vanishing gradient problem and therefore are able to capture long-term dependencies in time series (Hochreiter and Schmidhuber 1997).

Attention based models do not process time series sequentially and therefore are suitable to better parallelize the learning process. The time dependencies between individual time steps are learned from scratch. To make this easier, positional encodings are added to the individual states in this study, which provide information about the relative position of the state in the time series. To calculate the masked dot product attention matrix, the matrices $Q, K, V \in \mathbb{R}^{T,d}$ (query, key, value) and the mask $M \in \mathbb{R}^{T,T}$ are required as input according to Figure 6. In the case of self-attention Q, K, V are the same. The mask shown in Figure 6 is a look ahead mask. The masked (black) cells contain high negative values and are added to the scaled result of the matrix multiplication of Q and K. The subsequent use of the softmax function prevents to put attention on dependencies between already observed and future states. The Softmax function transforms a T-dimensional vector with real components into a T-dimensional vector $\sigma(z)$ also as a vector of real components in the value range [0, 1], where the components add up to 1.

$$\sigma(z)_t = \frac{e^{z_t}}{\sum_{t=0}^{T} e^{z_t}} \quad t = 1, \dots, T \tag{1}$$

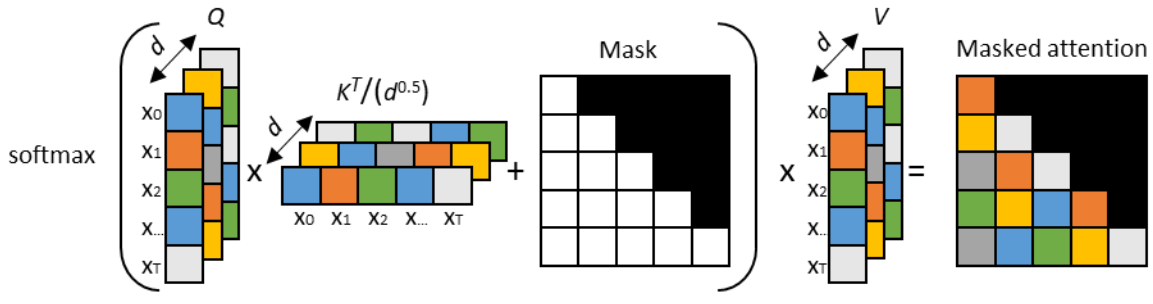


Figure 6: Illustration of the masked scaled dot product self-attention mechanism of an autoregressive model based on (Vaswani et al. 2017)

Before the dependencies between individual states can be learned in the LSTM/attention layers, layers must be introduced that use all the available information of a single state as input and learn its state representation in a multidimensional space.

Figure 7 b./c. show the different kinds of input provided to the autoregressive and imputation models and their first layers. Input to the autoregressive model is provided in the form of the mobility state $ms_{p,t}$, the time of the day/week τ_t, the day of the week d_t of person p at timestep t and as socio-demographic information $sd_{i,p}$. The time of the day/week is translated into a sinusoidal positional encoding using periods of one day/week. This is a typical approach to provide information about cyclical characteristics in time series (e.g. daily/weekly patterns) to the model. All other model inputs ($ms_{p,t}$, d_t, $sd_{i,p}$) are categorical and are therefore inserted into an embedding layer. Through the embedding layer the categorical information is mapped into a m-dimensional continuous space. The weights of the embedding layer and therefore the way the categorical variables are represented in the m-dimensional space are learned during the training process of the model. Further on, all the time step specific information are concatenated. The input time series is shifted one time step to the right ($t = 0 \dots T-1$) and starts with a dummy time step at $t = 0$, which is composed of a start token consisting of the start time and day and socio demographic information of the specific person. This training method is called teacher forcing (Williams and Zipser 1989).

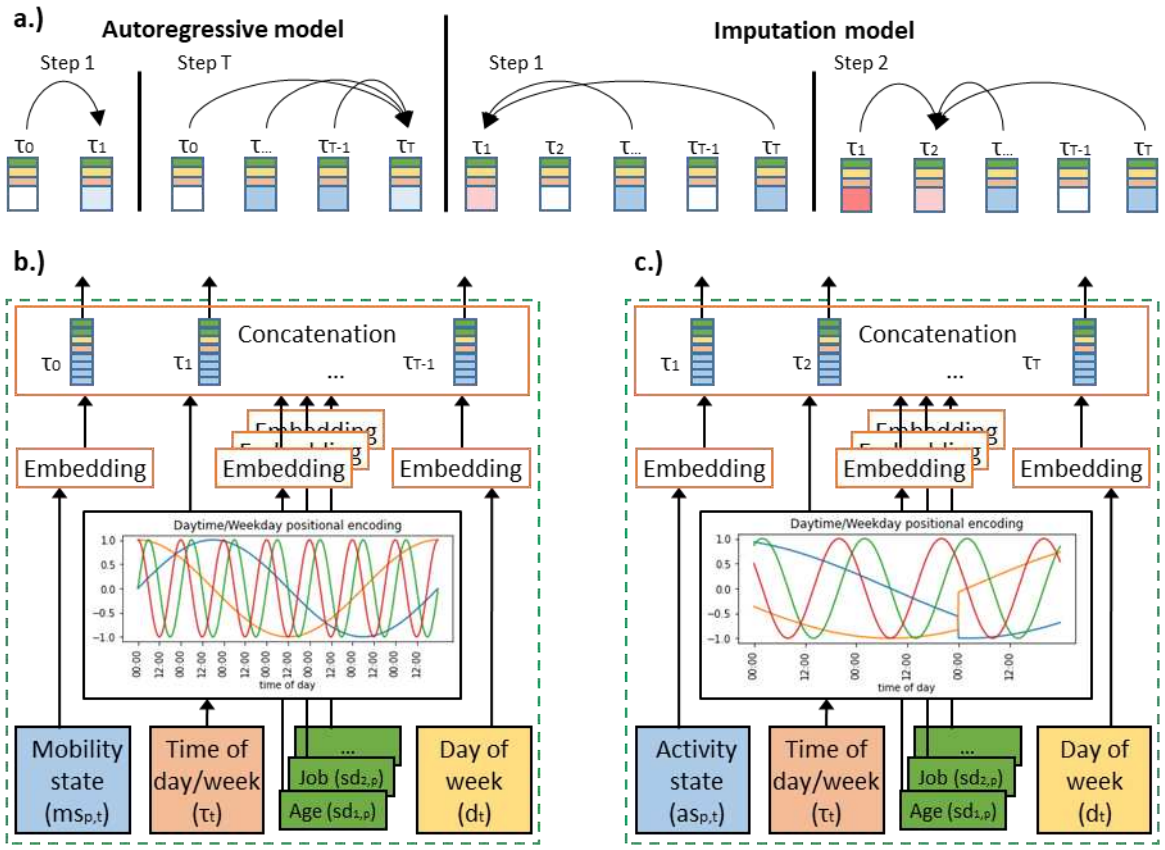


Figure 7: a.) Illustration of the relevant time step specific dependencies in the autoregressive and imputation models, b./c.) training input of the autoregressive/imputation (b./c.) models and visualization of their first layers

Figure 8 a. describes the central components of the LSTM based autoregressive model. After concatenating the time specific information, the vector state representations are fed into a linear dense layer before the state representations are inserted into a sequence to sequence LSTM layer. The final dense layer contains $|ms| = 6$ neurons which represent the probabilities (logits) of each mobility state $ms_{p,t}$ $(t = 1 \dots T)$.

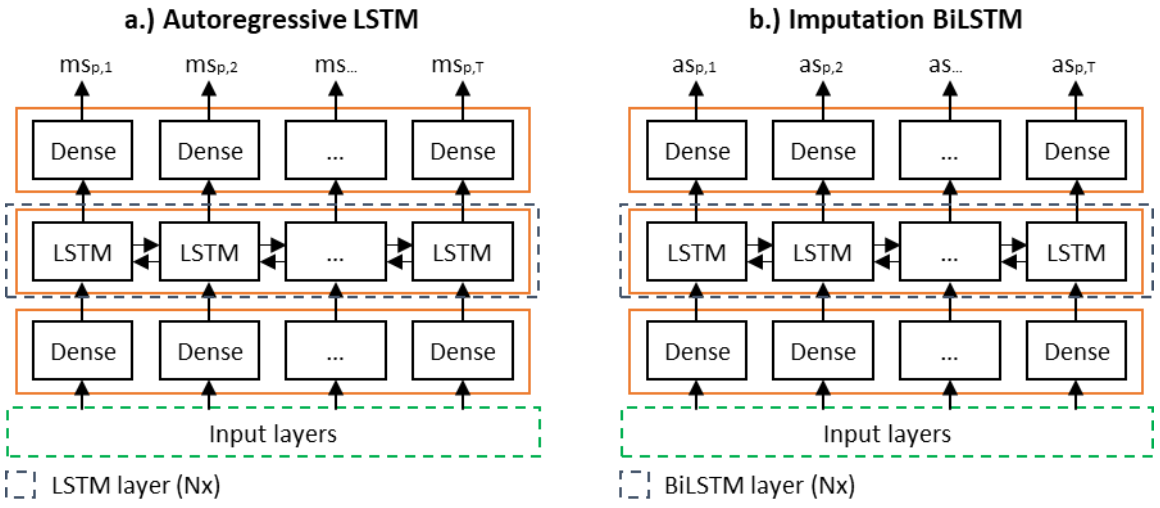


Figure 8: a.) LSTM based autoregressive model architecture and b.) BiLSTM based imputation model architecture

Figure 9 describes the architecture of the attention based transformer model. The transformer layer consists out of three linear dense layers for Q, K, V, the attention layer consisting of the scaled dot - product attention and two feed forward dense layers with dropout similar to (Vaswani et al. 2017).

Both models are trained by minimizing the cross entropy loss between the ground truth and the predicted probabilities.

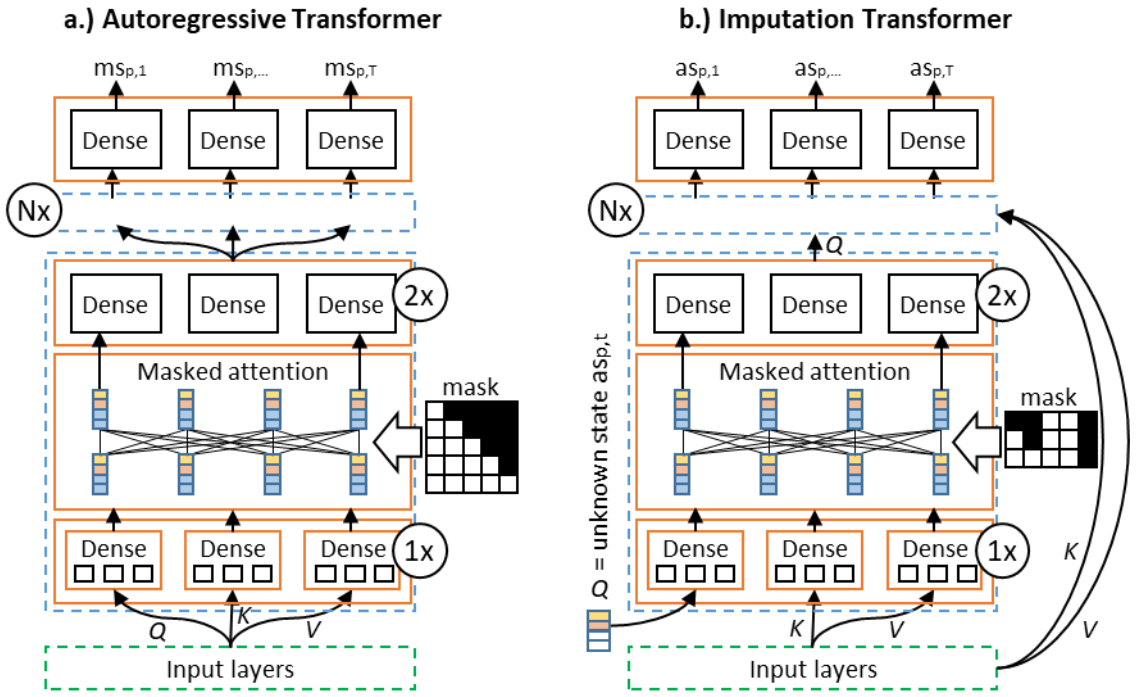


Figure 9: a.) Transformer based autoregressive model architecture and b.) Transformer based imputation model architecture (residual connections are not visualized)

3.2.2. Energy related activity imputation / enrichment

In the second model step, the generated weekly mobility plans are enriched with energy-related activities. A bidirectional LSTM model (Figure 8 b.) is compared with an attention-based transformer model (Figure 9 b.). In contrast to the first model step, information about individual mobility behavior over the entire week is already available when the first "at home" activity is estimated, this information has an impact on the activity choice. The procedure of the prediction process of the imputation model can be found in Figure 7 a.

As input data during the training process, the model is provided with activity time series of individual persons over 3 days (2x weekday, 1x weekend), the time and day of the week as well as socio-demographic parameters (job, age). The time step specific input processing can be seen in Figure 7 c. In contrast to the autoregressive models, the imputation models do not necessarily receive consecutive days as input, as this is not possible due to the structure of the time use survey. The connection between the three respective days is learned in the training process and applied to a whole week in the imputation process. In contrast to Figure 8 a., it can be seen in Figure 8 b. that the bidirectional LSTM architecture also takes future states into account when predicting the current state.

In contrast to the autoregressive transformer model, the imputation transformer does not use self-attention. The query vector Q of the first transformer layer contains the information about the unknown home states (unknown state, time, day, socio-demographic information). The key and value vector are identical and contain information about the mobility states of the three days (during training) or the week (during prediction). During the training process, *at home* activities of the TUD are masked and fed to the model as input. In all of the following Transformer layers, the output of the previous Transformer layer represents the query vector Q. The imputation models are trained using the cross entropy loss function.

3.3. Metrics

To evaluate the models presented, metrics must be introduced on the basis of which the model output can be assessed on an individual and aggregated level. The metrics presented below are generated and visualized at constant intervals during the training process.

The model-specific metrics are the cross entropy loss, which is minimized during the training process, and the model accuracy which provides information about how well the model predicts the next state. For the evaluation of the generated activity schedules, metrics are used to assess whether the proposed models reflect the variability in human behavior. Furthermore, metrics describing the variability of intrapersonal behavior are used to assess the consistency within a person's activity plan.

The aggregated state probability (sp) describes the aggregated probability $sp_{s,t}$ of a state $s \in S$ at time step $t \in T$ over a sample with the sample size N.

$$sp_{s,t} = \frac{\sum_{i=1}^{N} x_{i,s,t}}{N} \quad \forall\, s \in S, t \in T \tag{2}$$

State durations (sd) are calculated for all states $s \in S$ and are visualized by their cumulative distributions. The distribution of the duration of states can be used as a first indicator to evaluate the models with regard to the consideration of long-term time dependencies. For the evaluation of the intrapersonal variability within an activity schedule, the number of activities per week (na), the autocorrelation (ac) and the Hamming distance (hd) are calculated for each activity schedule of a sample. The autocorrelation is calculated for each activity state and each individual and is used to obtain information about the regularity of activities. The Hamming distance is calculated between all working days $d \in \{1 \dots 5\}$ of the week and thus provides information about the similarity of the daily behavior of individuals.

$$hd_n = \sum_{d_1=1}^{5} \sum_{d_2=1}^{5} |\{t \in \{1, \dots, T_d\} |\, s_{d_1,t} \neq s_{d_2,t}\}| \quad \forall\, n \in N \tag{3}$$

From the variability of these metrics (na, ac, hd), information about the diversity in behavior can be obtained.

4. Results

The results presented below were calculated with an XLA compiler and a "Tesla V100-SXM2-16GB" GPU in Tensorflow 2.3. To provide the models from overfitting, the data sets are randomly split up into training data (9-fold cross validation → 80 % training, 10 % validation) and test data (10 %).

4.1. Mobility schedule generation

As a reference model for the presented autoregressive models, ***a time-inhomogeneous*** first order Markov model is used. The first order Markov model characteristics are representative for the models presented in Section 2.1, since marginal changes in the metrics can be achieved by using more complex Markov chains, but the basic problems remain (no long-term memory). The introduced metrics are visualized in Figure 10. All metrics shown are calculated based on a sample size of N = 2,000 unless explicitly stated otherwise. The course of the aggregated state probability of the state *outside* deviates only slightly from the empirical course. The averaged root mean square error (rmse) over all states of the aggregated state probability is 0.53 % and tends towards zero with increasing sample size. From the course of the cumulative state durations of the state *mobile (car driver)* and the other states shown in Figure 16 it can be observed that the state durations of the schedules produced by the first order Markov model partly deviate from the empirical data. Furthermore, the distribution of the Hamming distance and the autocorrelation clearly differ between the data generated by the Markov model and the empirical data, which is reflected in large deviations in the rmse of the autocorrelation and the

mean absolute error of the Hamming distance. The peak in the autocorrelation in mobility behavior after 144 lags (one day) describes daily mobility patterns in the mobility behavior of individual persons. This peak, which can be clearly identified in the empirical data, is not represented in the synthetic mobility schedules of the Markov model. Compared to the empirical distribution, the distribution of the Hamming distances is shifted to the right, towards higher distances. Consequently, subsequent days of single individuals differ more from one another than in the empirical data. The distribution of number of activities per week indicates that the Markov model matches the empirical data well on average, but the boxplot indicates that the diversity in behavior deviates from the one observed in the empirical data.

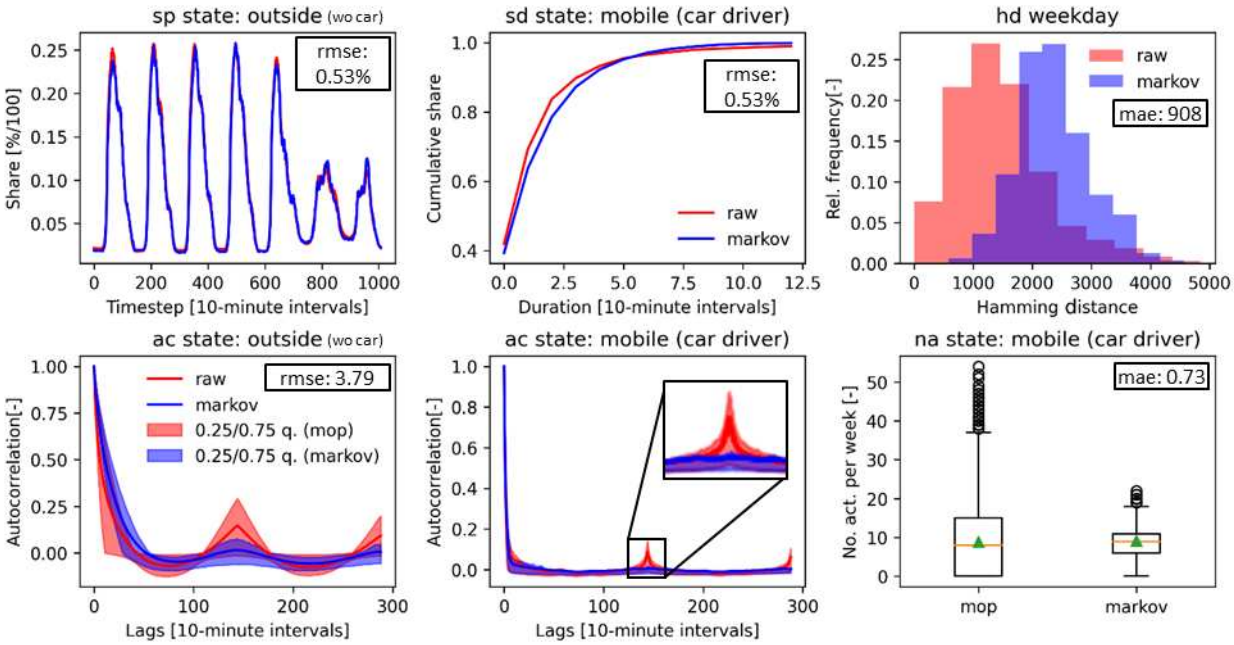


Figure 10: Visualization of the metrics for empirical MOP data (N = 26,610) and data generated with a first order Markov model (N = 2,000) (blue). The shown state dependent errors are calculated over all states and the mean is presented.

The autoregressive models presented in Section 3.2.1 are trained to predict the multinomial state distribution of the subsequent state. To achieve this, the cross entropy loss is minimized. Figure 11 and Figure 12 describe the course of the cross entropy loss during the training process. An epoch is defined as one training step of the nine-fold cross validation. After nine epochs, the training and validation data set are reshuffled and divided into nine new participations. During the training of the attention-based models, the loss function converges continuously for the training and test dataset. In the LSTM-based model, however, it can be seen that the course of the loss and accuracy function of the test data set diverges from the course of the training and validation data after around 14 epochs. From this point on, the model overfits on the training data and the training process can be stopped. In order not to use over-trained models, the weights of the model are saved at constant intervals during the training process. Furthermore, the development of the model accuracy during the training process is shown. This converges to a value of approx. 96.3%. This means that 96.3% of the time the correct value is predicted in the training process. Of course, the prediction is easier during the night when people are asleep than, for example, in the afternoon when there are many changes in activity. Figure 11 shows the course of the cross entropy loss for two model configurations, with one transformer layer and with four transformer layers. By increasing the depth of the neural network, the model can better map the complexity of mobility behavior. However, only marginal improvements can be achieved by further increasing the number of transformer layers from four to eight (Table 3). Since the performance of the models presented depends heavily on the choice of hyperparameters, various parameter settings were tested during the training phase for the LSTM and the attention based models. The parameter settings

varied during the training process and the corresponding metrics can be found in Table 2 and Table 3. ***In addition to the learning rate and the batch size, the number of LSTM units was varied, which limits the complexity of the internal state of the LSTM and is therefore important to capture temporal dependencies in behavior. The number of dense neurons (LSTM) or the model dimension (transformer) was varied to ensure that state-specific information is appropriately represented. Furthermore, the depth of the neural networks was varied, as this enables the neural network to learn higher level representations in human behavior.*** The results of the parameter variations show that the attention-based models are slightly superior to the LSTM-based models in most metrics, consequently, the attention-based model no. 3 from Table 3 is used for the presentation of the mobility schedule specific metrics.

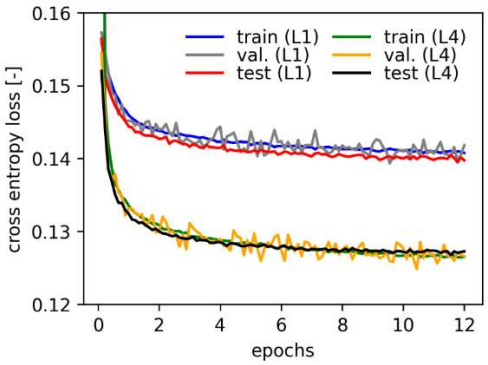


Figure 11: Loss development during training of the autoregressive transformer (L1/L4: 1/4 transformer layers)

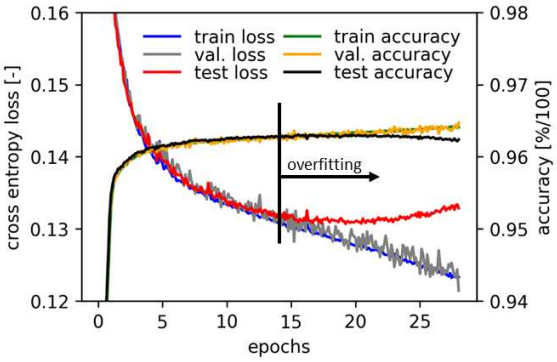


Figure 12: Loss and accuracy development during LSTM training

Selected mobility schedule specific metrics for the attention based autoregressive model described in Table 3 (model no. 3) are presented in Figure 13. A holistic overview of all metrics for all states can be found in the appendix (Figure 16). In contrast to the first order Markov model, the aggregated state probability is represented slightly worse by the attention based model. The rmse of the state probability averaged over all states and time steps is higher than the error of the first-order Markov model for all the models shown in Table 2 and Table 3 in the appendix. The Markov error corresponds to the standard error that arises with a sample size of 2,000. The standard error was calculated by randomly sampling 2,000 samples 30 times from the entire population and calculating their deviation from the metrics of the entire population (N = 26,610). The mean value of the error of the 30 samples is called the standard error. The mean absolute error of the number of weekly activities in the attention-based model is also higher than that of the Markov model (3.6 > 0.73). The diversity of the number of weekly activities is, however, recorded much more accurately by the attention-based model, which is shown in the lower right illustration in Figure 13 for the state *mobile (car driver)* and in Figure 16 for all other states. While the machine learning models presented in this work have slight deviations in the description of the averaged behavior and therefore perform slightly less accurately than Markov models, the mobility schedules generated differ fundamentally on the individual level, which is shown by the distribution of the cumulative state durations, the Hamming distance between weekdays and the autocorrelation of the individual states. Using the Hamming distance and the autocorrelation, it can be clearly seen that day-to-day dependencies in behavior are very accurately taken into account by the models presented in this work. In order to be able to adequately capture daily rhythms in mobility behavior, it is very important that the peak in the autocorrelation graph is captured well after 24 hours (144 10-minute time steps), which can be seen in the bottom center graph in Figure 13. From the course of the mean values and the ranges of the 25% / 75% quantile, it becomes clear that both these dependencies in the mean and in the spread are well represented across the entire population. These visual findings are also reflected in the significantly lower rmse of the autocorrelation compared to the Markov model (0.54 <3.79).

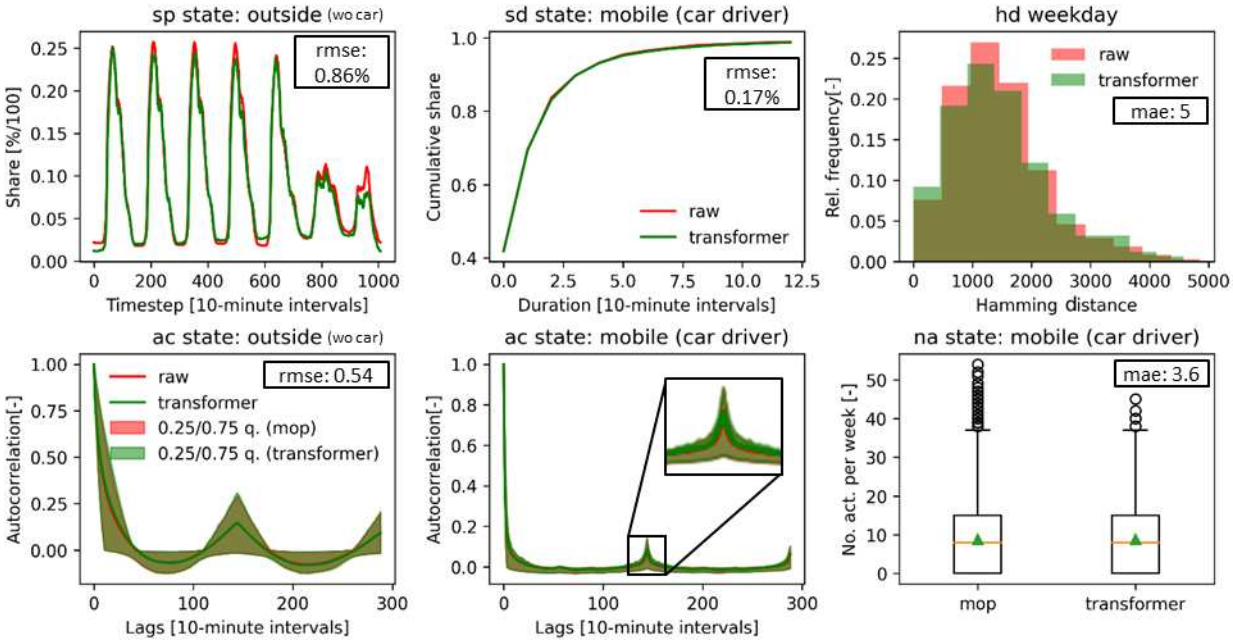


Figure 13: Visualization of the metrics for empirical MOP data (N = 26,610) and data generated with an attention based model (N = 2,000) (green). Model parameters can be seen in Table 3 (no. 3). The shown state dependent errors are calculated over all states and the mean is presented. The overlapping green and red ranges in the left-bottom and center-bottom graph describe the 25%/75% quantiles.

The difference between LSTM-based models and attention-based models is particularly evident from the autocorrelation peak in mobility behavior after 24 hours. LSTM models are also able to recognize relationships over such long periods of time, but in this work it was not possible to reproduce the peak as well with LSTM-based models as it can be seen in Figure 13 (bottom center) with the attention-based model. In addition to the low deviation of the mean error in the distribution of the Hamming distance (5 < 908), it can also be clearly recognized from the form of the distribution that the diversity in the profiles generated matches the real distributions much better than that of the Markov models, in which individual weekdays of a person do not have the similarities found in the empirical data.

4.2. Energy-related activity imputation

Since the model approach presented in this paper (step-by-step simulation of mobility behavior and subsequent enrichment of the results with energy-related activities based on different data sets) is new and no classical comparable applications in the field of behavioral modeling are known, only the results of the imputation models presented in Section 3.2.2 are benchmarked against each other in this section. As with the autoregressive models, the model performance of the imputation models is strongly dependent on the choice of hyperparameters. The parameters of the BiLSTM-based and the attention-based imputation model that were varied during the training process can be found in Table 4 and Table 5 in the appendix. To ensure that dependencies between time steps can be adequately captured by the model, sufficient amounts of LSTM units and attention layers must be provided. The dimension of the model must be chosen so that all time-step-specific information can be mapped well. In the following, the activity schedule-specific metrics for the attention-based model no. 6 from Table 5 are compared with the empirically collected TUD data. The metrics are visualized for specific states in Figure 14. A holistic overview of all metrics for all states and the development of the model loss and accuracy can be found in the appendix (Figure 17/Figure 18).

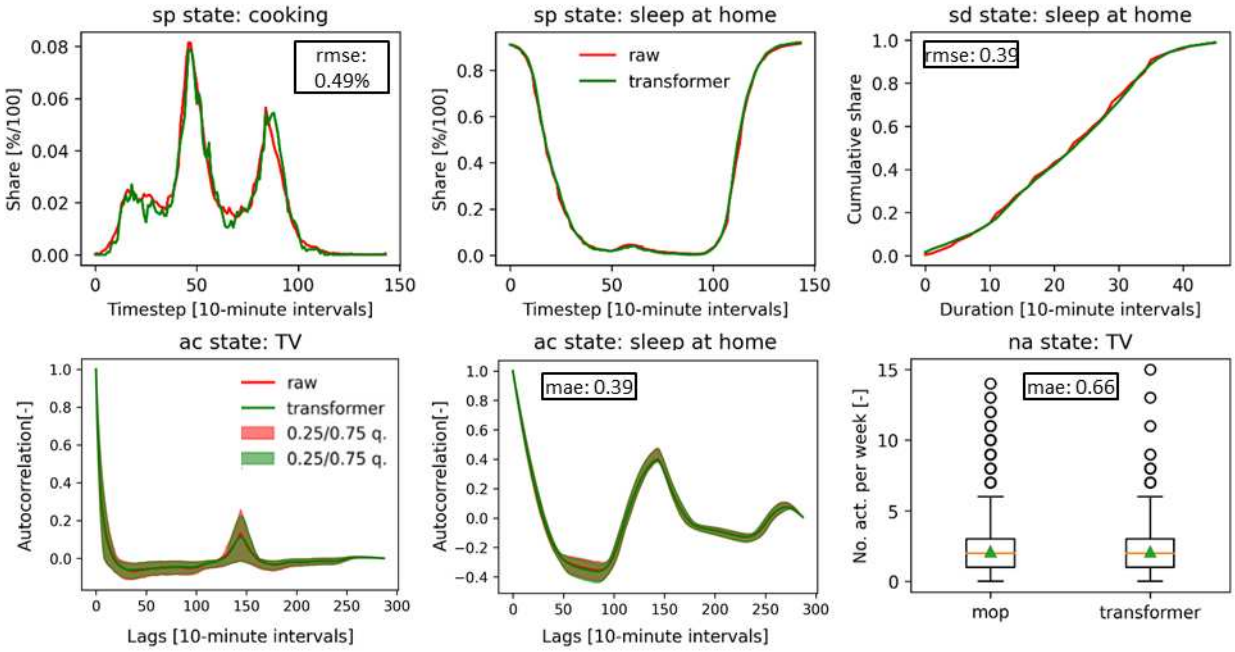


Figure 14: Visualization of the metrics for empirical TUD data (N = 35,691 dairy days) and data generated with an attention based model (N = 2,000 diary days) (green). Model parameters can be seen in Table 5 (model no. 6). The shown state dependent errors are calculated over all states and the mean is presented. The overlapping green and red ranges in the left-bottom and center-bottom graph describe the 25%/75% quantiles. The autocorrelation graphs were calculated based on the two work days over 288 10-minute timesteps.

Similar to the autoregressive models, it can be seen from the course and the rmse of the aggregated state probability that this differs slightly from the empirically collected data. The averaged errors over all states and time steps can be taken from Figure 14, Table 4 and Table 5 for the various model variants. The error in the simulation of the state durations, on the other hand, is smaller than that which occurs when modelling activities with a first-order Markov chain (no imputation model). Since the German TUD data set contains diary entries for three days of the week, the model can also learn day-to-day dependencies between energy-relevant activities. The autocorrelation graphs in Figure 18 show that the imputation model is able to recognize and reproduce these dependencies. For example, daily sleep rhythms can be reproduced in the synthetic data, which is another unique selling point of this work.

When comparing the metrics shown in Table 4 and Table 5, it is noticeable that the attention-based models perform slightly better in representing the aggregated state probability, while the BiLSTM-based models tend to map the duration of states and autocorrelation better. This could be attributed to the fact that when representing energy-relevant activities, short-term temporal dependencies between individual states are of higher importance than the one seen in the mobility schedules and the sequential character of the BiLSTM depicts these dependencies well, while attention-based models tend to capture individual states and their time-dependent probability of occurrence more strongly than short-term sequential dependencies.

4.3. Generation of weekly activity schedules

After the training processes of the autoregressive models and the imputation models have been described and evaluated in Sections 4.1 and 4.2, synthetic weekly activity plans are now generated for various socio-demographic groups and compared with empirical data. Table 6 in the appendix gives an overview of the socio-demographic composition of the empirical data. The age distribution of the MOP data shows that older population groups are overrepresented in contrast to the TUD data. Younger groups of the population such as students and part-time workers, on the other hand, are under-

represented. Due to the consideration of socio-demographic factors when coupling the data sets in the approach presented, a different distribution of the socio-demographic groups in the individual data sets is not problematic. When considering the sample sizes of the MOP and TUD data, it must be taken into account that the TUD samples, in contrast to the MOP samples, only consist of two to three days. The MOP data set with 10-minute time resolution has more than five times as many data points as the TUD data set. From the rmse of the aggregated state probabilities for the different socio-demographic groups, it can be seen that the data sets differ in some cases more strongly (rmse (age <18): 4.0%). In the synthetic profiles, the mobility behavior is generated on the basis of the MOP data, consequently, when looking at the rmse, fewer errors can be found between the synthetically generated data and the MOP data, both when looking at the socio-demographic groups in a differentiated manner and when looking at the aggregate as a whole dataset.

Finally, Figure 15 shows the course of the aggregated state probabilities over a week and two exemplary activity plans of synthetically generated schedules for two socio-demographic groups (age<18, full time employees). From the visualization of the aggregated state probabilities it can be seen that children under the age of 18 are mainly out of the home in the mornings and have two pronounced mobility peaks at around 8 am and 1 pm, while full-time employees are mainly outside during the day. Rhythmic behavior within the working days can be seen in the exemplary individual profiles. In the activity plan of the student on Friday morning, the student changes from an *at home* state to an *outside* state without a mobility activity in between. At first glance, this seems unrealistic, but these transitions can also be found in the empirical data due to the temporal aggregation of the mobility data over 10 minutes.

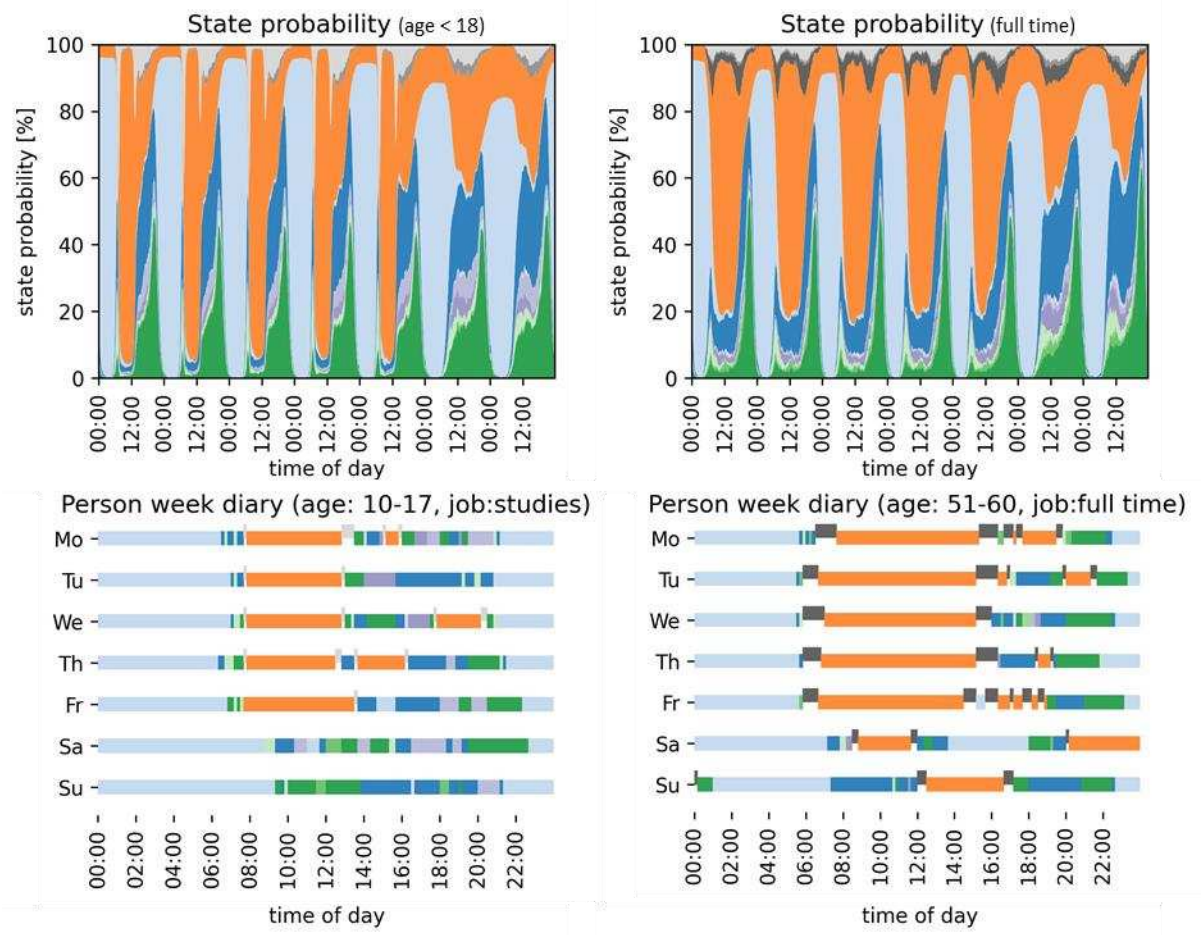


Figure 15: The top two figures represent the course of the aggregated state probability for 1,500 generated activity plans for persons under 18 years of age and for full time employees. The lower two representations are two exemplary activity plans for a person under the age of 18 and a full-time employee (A legend can be found in Figure 4).

5. Discussion

The results of Section 4.1 show that the Markov model used as a reference model is not able to record long-term dependencies in activity patterns and, due to the structure of the approach, is not able to adequately record the diversity in occupancy behavior. Consequently, synthetic activity schedules generated with Markov chains cannot be used to analyse occupancy behavior on an individual level and are only suitable for studies on an aggregated level. The approach presented in this paper combines weekly mobility data with a large sample size with high-resolution activity data with the help of new machine learning algorithms. The approach creates a new data basis which can be used for further analyses of home occupancy and mobility behavior. The profiles generated have similar stochastic properties as the empirically collected data on both the individual and the aggregated level.

By adequately capturing long-term dependencies in people's activities, the behavior of individual people can be reproduced. As a result, the data generated represent the basis for a variety of potential applications, one of which is the examination of potential charging periods of people with electric vehicles, assuming that electric mobility does not change mobility behavior. By combining the detailed mobility data with high-resolution activity data, a unique data basis is created which offers the possibility of consistently simulating the energy demand from personal mobility, the electrical demand for household devices and the heat demand for space heating and domestic hot water. Therefore, simultaneity effects in energy demand can be analysed based on one fundamental data set.

When analyzing such future developments, it should be taken into account that the data sets on which this work is based describe historical behavior (MOP: 2001-2017, TUD: 2001/02). Not taking into account the dynamics in people's behavioral habits could lead to significant errors, depending on the application. The energy sector includes many examples of innovations that have changed people's behavior for example, the internal combustion engine for transport and the development of ICT in recent decades. Hence ground-breaking/disruptive technologies could change the nature of the energy service demand itself (e.g. autonomous electric vehicles and smart home applications). In order to take into account temporal changes in behavior in the data set, the survey year of the respective sample could be provided as additional information in future studies. Furthermore, the data sets used differ in their temporal resolution, while the mobility data (MOP) are available in minute resolution, activities in the TUD are recorded in ten-minute resolution. The aggregation of the mobility data to a temporal resolution of 10 minutes can lead to distortions in short mobility states.

Through the use of machine learning approaches the assumption bias in the presented approach is low in comparison to e.g. utility-based stepwise regression approaches (Hilgert et al. 2017), therefore the developed approach is highly transferable. TUD data are collected uniformly in several European countries, but there are some differences in the design of the surveys. Some countries only provide activity time series for one weekday and one weekend day, which makes it harder to capture interday dependencies in activities. Longitudinal surveys of mobility behavior are not carried out in a harmonized way at the European level. However, similar mobility studies are available, for example in the UK and the Netherlands, which examine the mobility behavior over a whole week of a sample that is representative of the nation (Department for Transport 2020; Hoogendoorn-Lanser et al. 2015). The approach presented could therefore easily be applied to behavioral data in the UK and the Netherlands. Instead of training individual models for different countries, it would make more sense to implement the country information as a socio-demographic parameter in a transnational model in order to learn country-specific behavior and at the same time provide the model with a larger database for learning general behavioral relationships.

In this work, the focus was placed on the mapping of the mobility and activity behavior of individual persons and therefore no interpersonal relationships in the behavior of several individuals in a household were taken into account. However, the presented approach can and will be extended to represent household behavior in order to capture interpersonal relationships. Furthermore, only socio-demographic behavioral differences based on age and employment are currently taken into account in the model. Since the underlying data sets contain significantly more socio-demographic differentiations, an extension to include further socio-demographic characteristics is possible.

Since the training process is stopped before the presented models overfit, it can be stated that the models have learned the general stochastic relationships in human behavior and not simply learned the raw data sets by heart. This statement is supported by Figure 19 in the appendix, which describes the distribution of the minimum distances of a sample of data set a with all samples of data set b. The distribution of the minimum distances between the synthetic mobility schedules and the raw data is similar to the distribution of the minimum distances within the empirically collected data. However, even if the raw data used in this paper are already provided in anonymized form, it must be ensured that no information about individual samples in the empirical data is revealed by the synthetic data sets. Consequently, in follow-up work, prior to making the models presented in this paper available to the general public, algorithms from the field of "differential privacy" must be used to ensure that no information about individual samples is provided (Dwork and Roth 2014). Algorithms that ensure the privacy of individuals have been developed in recent years for deep learning applications (Abadi et al. 2016). Ensuring differential privacy is always accompanied by a loss of quality in the model, whereby this trade-off between quality and privacy can be clearly quantified by the so-called privacy budget.

6. Conclusion and Outlook

Over the past few years, many models have been published that aim to capture relationships in activity patterns to explain residential energy demand. Most of these models are different Markov variants or regression models that have a strong assumption bias and are therefore unable to capture complex long-term dependencies and the diversity in occupancy behavior. In this work it was shown that machine learning models from the field of natural language processing are able to capture long-term dependencies in mobility and activity patterns and at the same time adequately depict the diversity in behavior across the entire population. In a first step, two autoregressive models are presented which are able to recognize and reproduce weekly mobility patterns. In a second step, two imputation models are trained with time use data, which, based on the mobility information of individual people, enrich them with energy-related activities. Finally, the two models are combined to generate weekly activity plans. By combining an autoregressive generative model with an imputation model, the advantages of two data sets are combined and new data are generated which are beneficial for multiple use cases. One of which is the examination of flexibility potentials of individual households which is urgently needed for the integration of volatile renewable energy sources. Furthermore, metrics were introduced that enable activity profiles to be investigated in terms of intrapersonal and interpersonal variability. Based on these metrics, it is shown that the synthetically generated activity plans represent weekly mobility patterns and day-to-day dependencies of the energy-relevant activities with a high quality on an individual and aggregated level. The evaluation metrics show that LSTM and attention-based neural networks outperform existing approaches on an individual level by a large margin and at the same time have only slight deviations in the aggregated behavior.

Due to the availability of rich socio-demographic information in the two basic data sets, activity plans can be generated for different socio-demographic groups and can be used in future work to simulate consistent energy demand profiles from electric mobility, household devices and space heating. The approach developed provides the basis for making high-quality weekly activity data available to the general public without having to carry out complex application procedures. It was shown that the presented approach does not learn the training data by heart, however, it must be ensured that no private information about individuals is revealed by the model before the synthetic data can be provided to the community, which cannot be ensured at the current time. Therefore, in further work the model will be trained in a differential private way. Furthermore, the presented methodology can be trained with behavioral data from different European countries in order to develop a transnational model. Instead of individual behavior, household behavior could be learned to take interpersonal dependencies into account.

Acknowledgement

This work was supported by the Helmholtz Association under the Joint Initiative “Energy Systems Integration” (funding reference: ZT-0002) and was done during a research stay funded by the Centre for Research into Energy Demand Solutions (CREDS) at the University of Reading (UK). This work was supported by UKRI [grant numbers EP/R000735/1, EP/R035288/1 and EP/P000630/1].

7. References

Abadi, Martin; Chu, Andy; Goodfellow, Ian; McMahan, H. Brendan; Mironov, Ilya; Talwar, Kunal; Zhang, Li (2016): Deep Learning with Differential Privacy. In *CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318. DOI: 10.1145/2976749.2978318.

Aerts, D.; Minnen, J.; Glorieux, I.; Wouters, I.; Descamps, F. (2014): A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. In *Building and Environment* 75, pp. 67–78. DOI: 10.1016/j.buildenv.2014.01.021.

Bahdanau, Dzmitry; Cho, Kyunghyun; Bengio, Yoshua (2015): Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of International Conference on Learning Representations*. Available online at http://arxiv.org/pdf/1409.0473v7.

Bengio, Yoshua; Ducharme, Réjean; Vincent, Pascal; Jauvin, Christian (2003): A neural probabilistic language model. In *Journal of Machine Learning Research (3)*, pp. 1137–1155.

Bottaccioli, Lorenzo; Di Cataldo, Santa; Acquaviva, Andrea; Patti, Edoardo (2019): Realistic Multi-Scale Modeling of Household Electricity Behaviors. In *IEEE Access* 7, pp. 2467–2489. DOI: 10.1109/ACCESS.2018.2886201.

Bowman, John (1998): The Day Activity Schedule Approach to Travel Demand Analysis. Dissertation. Cambridge, Massachusetts.

Brown, Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla et al. (2020): Language Models are Few-Shot Learners, 5/28/2020. Available online at http://arxiv.org/pdf/2005.14165v4.

Caccia, Massimo; Caccia, Lucas; Fedus, William; Larochelle, Hugo; Pineau, Joelle; Charlin, Laurent (2020): Language GANs Falling Short. In *Proceedings of the Seventh International Conference on Learning Representation*. Available online at https://openreview.net/pdf?id=r1lOgyrKDS.

Chen, Zhenghua; Jiang, Chaoyang; Xie, Lihua (2018): Building occupancy estimation and detection. A review. In *Energy and Buildings* 169, pp. 260–270. DOI: 10.1016/j.enbuild.2018.03.084.

Department for Transport (2020): 2019 National Travel Survey.

Destatis (2006): Zeitbudgeterhebung: Aktivitäten in Stunden und Minuten nach Geschlecht, Alter und Haushaltstyp. Zeitbudgets - Tabellenband I. 2001/2002. Wiesbaden. Available online at https://www.statistischebibliothek.de/mir/receive/DEMonografie_mods_00003054.

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (2018): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Available online at http://arxiv.org/pdf/1810.04805v2.

Dwork, Cynthia; Roth, Aaron (2014): The Algorithmic Foundations of Differential Privacy. In *FNT in Theoretical Computer Science* 9 (3-4), pp. 211–407. DOI: 10.1561/0400000042.

Elman, Jeffrey L. (1990): Finding Structure in Time. In *Cognitive Science* 1990 (14), pp. 179–211.

Eurostat (2000): Harmonized European Time of Use Survey.

Fischer, David; Härtl, Andreas; Wille-Haussmann, Bernhard (2015): Model for electric load profiles with high time resolution for German households. In *Energy and Buildings* 92, pp. 170–179. DOI: 10.1016/j.enbuild.2015.01.058.

Flett, Graeme; Kelly, Nick (2016): An occupant-differentiated, higher-order Markov Chain method for prediction of domestic occupancy. In *Energy and Buildings* 125, pp. 219–230. DOI: 10.1016/j.enbuild.2016.05.015.

Goodfellow, Ian; Jean Pouget-Abadie; Mehdi Mirza; Bing Xu; David Warde-Farley; Sherjil Ozair et al. (2014): Generative Adversarial Nets. In *Advances in neural information processing systems, pp 2672–2680*.

Hilgert, Tim; Heilig, Michael; Kagerbauer, Martin; Vortisch, Peter (2017): Modeling Week Activity Schedules for Travel Demand Models. In *Transportation Research Record* 2666 (1), pp. 69–77. DOI: 10.3141/2666-08.

Hochreiter, S.; Schmidhuber, J. (1997): Long Short-Term Memory. Available online at https://www.mitpressjournals.org/doi/10.1162/neco.1997.9.8.1735, checked on 8/13/2020.136Z.

Hoogendoorn-Lanser, Sascha; Schaap, Nina T.W.; OldeKalter, Marie-José (2015): The Netherlands Mobility Panel: An Innovative Design Approach for Web-based Longitudinal Travel Data Collection. In *Transportation Research Procedia* 11, pp. 311–329. DOI: 10.1016/j.trpro.2015.12.027.

Kalchbrenner, Nal; Grefenstette, Edward; Blunsom, Phil (2014): A Convolutional Neural Network for Modelling Sentences. Available online at http://arxiv.org/pdf/1404.2188v1.

Kaschub, Thomas (2017): Batteriespeicher in Haushalten unter Berücksichtigung von Photovoltaik, Elektrofahrzeugen und Nachfragesteuerung. Dissertation.

McKenna, Eoghan; Krawczynski, Michal; Thomson, Murray (2015): Four-state domestic building occupancy model for energy demand simulations. In *Energy and Buildings* 96, pp. 30–39. DOI: 10.1016/j.enbuild.2015.03.013.

Mikolov, Tomas; Chen, Kai; Corrado, Greg; Dean, Jeffrey (2013): Efficient Estimation of Word Representations in Vector Space. Available online at http://arxiv.org/pdf/1301.3781v3.

Paardekooper, Susana; Lund, Rasmus Søgaard; Mathiesen, Brian Vad; Chang, Miguel (2018): Heat Roadmap Europe 4. Quantifying the Impact of Low-Carbon Heating and Cooling Roadmaps.

Pflugradt, Noah Daniel (2016): Modellierung von Wasser und Energieverbräuchen in Haushalten. Available online at http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-209036, checked on 4/12/2018.

Ramírez-Mendiola, José Luis; Grünewald, Philipp; Eyre, Nick (2019): Residential activity pattern modelling through stochastic chains of variable memory length. In *Applied Energy* 237, pp. 417–430. DOI: 10.1016/j.apenergy.2019.01.019.

Richardson, Ian; Thomson, Murray; Infield, David (2008): A high-resolution domestic building occupancy model for energy demand simulations. In *Energy and Buildings* 40 (8), pp. 1560–1566. DOI: 10.1016/j.enbuild.2008.02.006.

Richardson, Ian; Thomson, Murray; Infield, David; Clifford, Conor (2010): Domestic electricity use. A high-resolution energy demand model. In *Energy and Buildings* 42 (10), pp. 1878–1887. DOI: 10.1016/j.enbuild.2010.05.023.

Shove, Elizabeth; Pantzar, Mika; Watson, Matt (2012): The Dynamics of Social Practice: Everyday Life and How it Changes: SAGE Publications Ltd.

Steemers, Koen; Yun, Geun Young (2009): Household energy consumption. A study of the role of occupants. In *Building Research & Information* 37 (5-6), pp. 625–637. DOI: 10.1080/09613210903186661.

Sutskever, Ilya (2013): Training Recurrent Neural Networks. Dissertation.

Sutskever, Ilya; Vinyals, Oriol; Le V, Quoc (2014): Sequence to Sequence Learning with Neural Networks. Available online at http://arxiv.org/pdf/1409.3215v3.

Tjaden, Tjarko; Bergner, Joseph; Weniger, Johannes; Quaschning, Volker (2015): Repräsentative elektrische Lastprofile für Wohngebäude in Deutschland auf 1-sekündiger Datenbasis.

Torriti, Jacopo (2014): A review of time use models of residential electricity demand. In *Renewable and Sustainable Energy Reviews* 37, pp. 265–272. DOI: 10.1016/j.rser.2014.05.034.

826 Torriti, Jacopo (2017): Understanding the timing of energy demand through time use data. Time of the
827 day dependence of social practices. In *Energy Research & Social Science* 25, pp. 37–47. DOI:
828 10.1016/j.erss.2016.12.004.

829 Transport & Environment (2020): Recharge EU. How many charge points will Europe and its member
830 states need in the 2020s. With assistance of William Todts. Available online at
831 https://www.transportenvironment.org/sites/te/files/publications/01%202020%20Draft%20TE%20In
832 frastructure%20Report%20Final.pdf.

833 Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N. et al.
834 (2017): Attention Is All You Need. Available online at http://arxiv.org/pdf/1706.03762v5.

835 Walker, Gordon (2014): The dynamics of energy demand: Change, rhythm and synchronicity. In *Energy*
836 *Research & Social Science* 1, pp. 49–55. DOI: 10.1016/j.erss.2014.03.012.

837 Weiß, Christine; Chlond, Bastian; Hilgert, Tim; Vortisch, Peter (2016): Deutsches Mobilitätspanel (MOP)
838 - wissenschaftliche Begleitung und Auswertungen, Bericht 2014/2015. Alltagsmobilität und
839 Fahrleistung, checked on 10/22/2019.

840 Wilke, Urs (2013): Probabilistic Bottom-up Modelling of Occupancy and Activities to Predict Electricity
841 Demand in Residential Buildings. PHD Thesis.

842 Williams, R. J.; Zipser, D. (1989): A learning algorithm for continually running fully recurrent neural
843 networks. In *Neural computation* (1(2)), 270–280.

844 Wolf, Thomas; Debut, Lysandre; Sanh, Victor; Chaumond, Julien; Delangue, Clement; Moi, Anthony et
845 al. (2020): HuggingFace's Transformers: State-of-the-art Natural Language Processing. Available online
846 at http://arxiv.org/pdf/1910.03771v5.

847 Wu, Yonghui; Schuster, Mike; Chen, Zhifeng; Le V, Quoc; Norouzi, Mohammad; Macherey, Wolfgang
848 et al. (2016): Google's Neural Machine Translation System: Bridging the Gap between Human and
849 Machine Translation. Available online at http://arxiv.org/pdf/1609.08144v2.

850 Yamaguchi, Y.; Yilmaz, S.; Prakash, N.; Firth, S. K.; Shimoda, Y. (2018): A cross analysis of existing
851 methods for modelling household appliance use. In *Journal of Building Performance Simulation* (5673),
852 pp. 1–20. DOI: 10.1080/19401493.2018.1497087.

853 Yilmaz, Selin; Firth, Steven K.; Allinson, David (2017): Occupant behaviour modelling in domestic
854 buildings: the case of household electrical appliances. In *Journal of Building Performance Simulation*
855 10 (5-6), pp. 582–600. DOI: 10.1080/19401493.2017.1287775.

856 Young, Tom; Hazarika, Devamanyu; Poria, Soujanya; Cambria, Erik (2017): Recent Trends in Deep
857 Learning Based Natural Language Processing. Available online at http://arxiv.org/pdf/1708.02709v8.

858 Yu, Lantao; Zhang, Weinan; Wang, Jun; Yu, Yong (2017): SeqGAN: Sequence Generative Adversarial
859 Nets with Policy Gradient. Available online at https://www.nature.com/articles/nature16961.pdf.

860 Zumkeller, D.; Chlond, B. (2009): Dynamics of Change: Fifteen-Year German Mobility Panel. Presented
861 at 88th Annual Meeting of the Transportation Research Board. Washington, D.C., 2009.

862

8. Appendix

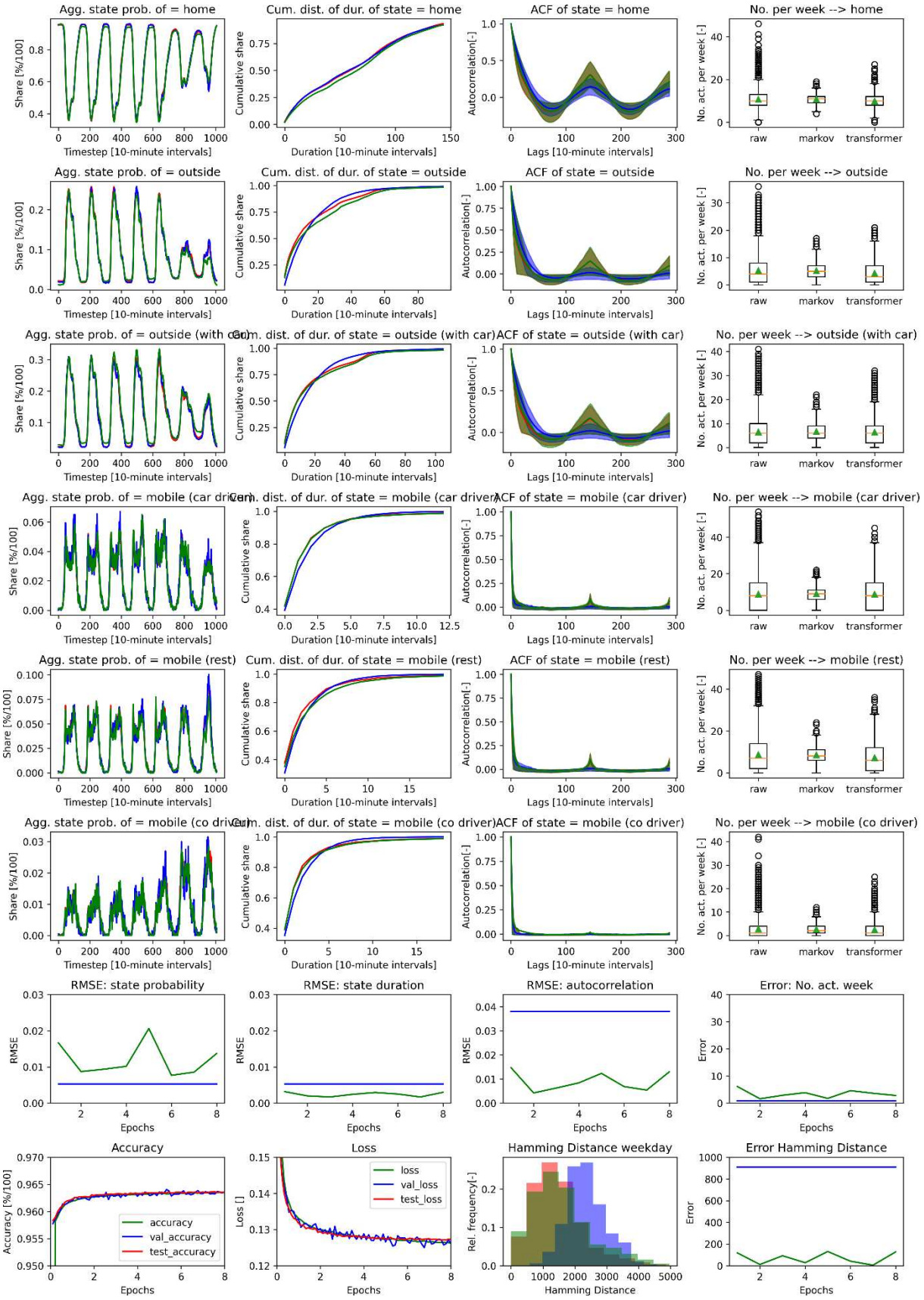


Figure 16: Comparison of all metrics and all states for the mop data (red), the attention based autoregressive model described in Table 3 (no. 3) (green) and a first order Markov model (blue). The mobility schedule specific metrics of the attention based model are calculated based on the model weights after epoch 7.

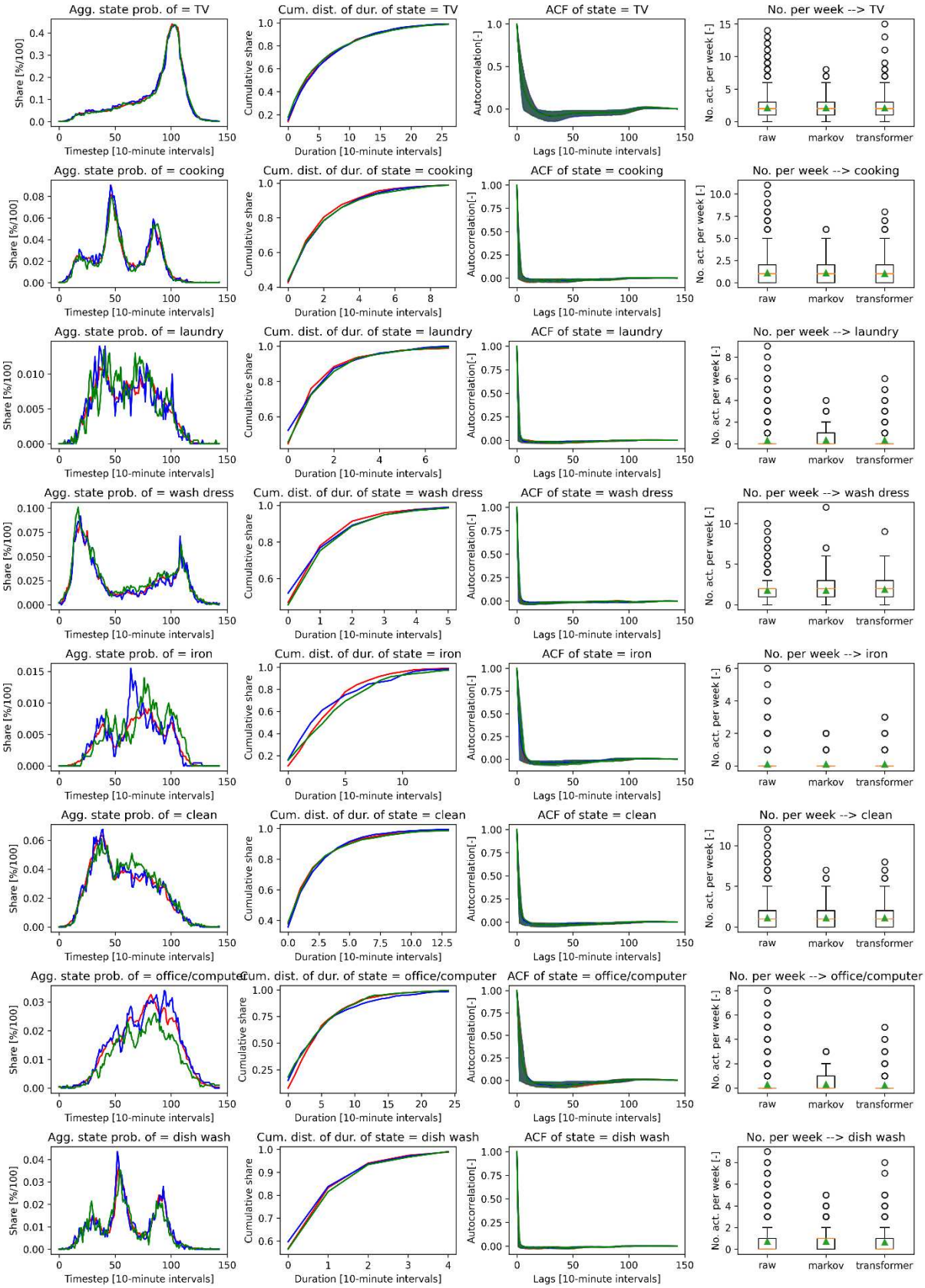


Figure 17: Part a: Comparison of all metrics and all states for the TUD data (red), the attention based imputation model described in Table 5 (model no. 6) (green) and a first order Markov model (blue – no imputation model). The mobility schedule specific metrics of the attention based model are calculated based on the model weights after epoch 37. The autocorrelation graphs were calculated based on single days.

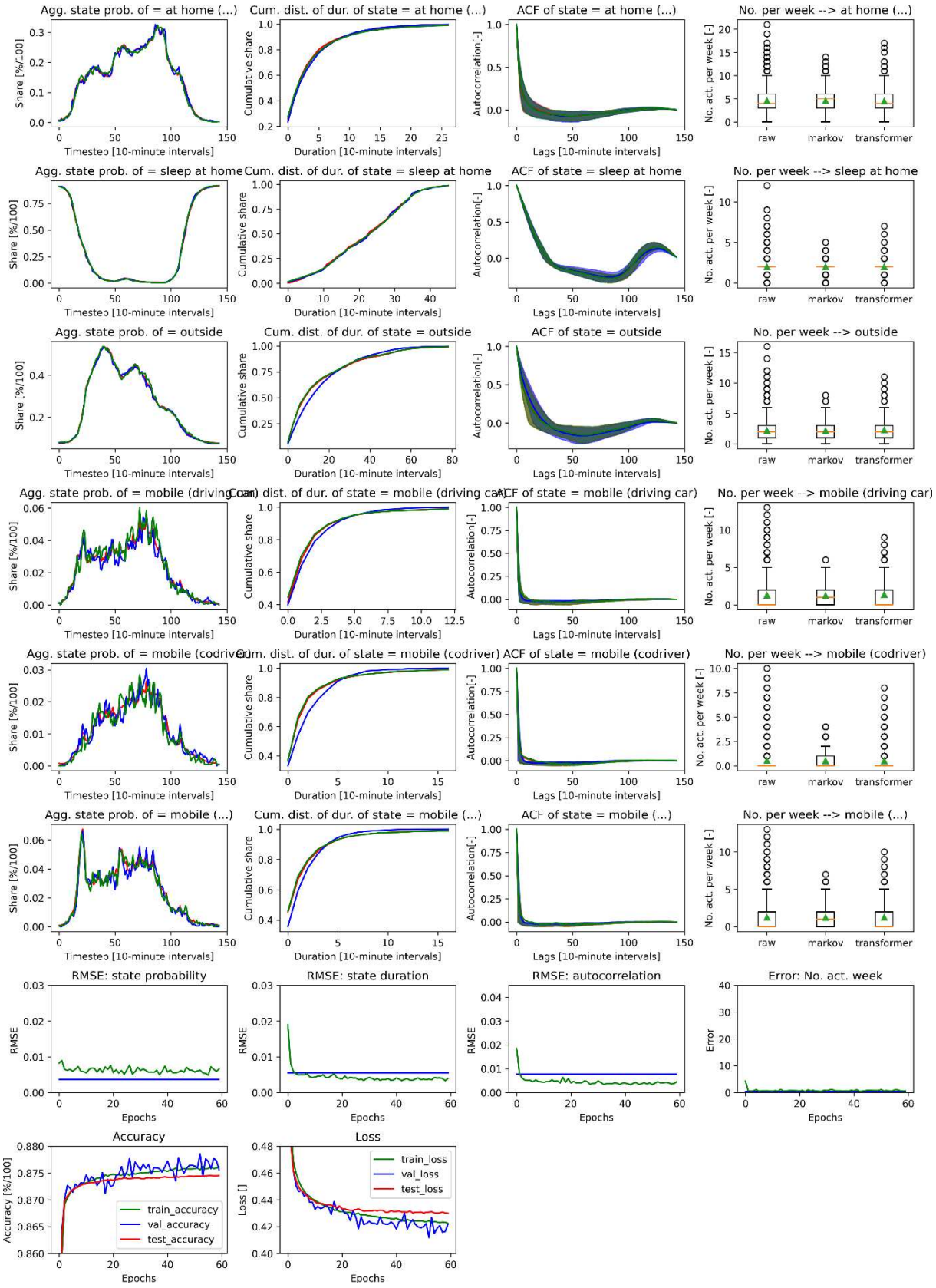


Figure 18: Part b: Comparison of all metrics and all states for the TUD data (red), the attention based imputation model described in Table 5 (model no. 6) (green) and a first order Markov model (blue – no imputation model). The mobility schedule specific metrics of the attention based model are calculated based on the model weights after epoch 37. The autocorrelation graphs were calculated based on single days. Furthermore, the course of the model loss and accuracy is visualized.

Table 2: Hyperparameter configurations and model metrics for the LSTM based autoregressive model. Metrics were calculated based on a sample size of N=2,000. Furthermore, a mean standard error due to the sample size of 2,000 is given.

No.	LSTM units/ Learning rate/ Batch size/ Dense neurons	Sp rmse [%]	Sd rmse [%]	Ac rmse [%]	Na mae []	Hd mae []	Cross-entropy Loss	Accuracy [%]	CV Epochs
1	512/0.0005/512/32	0.99	0.13	0.71	1.11	144	0.133	96.27	14
2	128/0.0005/512/32	1.03	0.17	1.65	1.57	423	0.142	96.12	8
3	512/0.001/512/32	1.05	0.18	0.66	3.04	235	0.131	96.30	11
4	512/0.0005/64/32	1.27	0.22	0.89	3.39	114	0.134	96.26	3
5	512/0.0005/512/64	0.90	0.18	0.80	0.67	98	0.131	96.29	17
6	512/0.0005/256/32	0.90	0.13	0.60	1.85	83	0.131	96.29	11
7	2x256/0.001/512/32	0.69	0.14	0.95	2.08	120	0.131	96.29	12
8	2x256/0.0005/256/32	0.97	0.19	0.63	3.61	1.5	0.132	96.28	10
Standard error (N=2,000)		0.52	0.09	0.24	0.6	13	-	-	-

Table 3: Hyperparameter configurations and model metrics for the attention based autoregressive model. 2xh means that two attention heads are used (see (Vaswani et al. 2017)).

No.	Transformer layers/ D_model/ Learning rate/ Batch size	Sp rmse	Sd rmse	Ac rmse	Na mae	Hd mae	Cross-entropy Loss	Accuracy	CV Epochs
1	1/64/0.001/64	0.83	0.31	1.32	2.96	244	0.14	95.95	9
2	4/64/0.001/64	0.91	0.16	0.70	2.53	33	0.128	96.34	15
3	**8/64/0.001/64**	**0.86**	**0.17**	**0.54**	**3.6**	**5**	**0.127**	**96.36**	**7**
4	4/64/0.001/128	0.95	0.22	0.54	3.28	44	0.130	96.29	3
5	4/128/0.001/128	0.89	0.24	0.59	3.60	9	0.128	96.33	6
6	4/64/0.0005/64	0.86	0.18	0.48	4.78	6	0.128	96.33	20
7	2(2xh)/64/0.001/64	0.97	0.22	0.60	6.33	74	0.129	96.31	11
8	4(2xh)/64/0.001/64	1.20	0.20	0.42	4.52	126	0.127	96.35	8
Standard errors (N=2,000)		0.52	0.09	0.24	0.6	13	-	-	-

Table 4: Hyperparameter configurations and model metrics for the BiLSTM based imputation model. Metrics were calculated based on a sample size of N=2,000 diary days. Furthermore, a mean standard error due to the sample size of 2,000 diary days is given.

No.	LSTM units/ D_model/ Learning rate/ Batch size	Sp rmse	Sd rmse	Ac rmse	Na mae	Cross-entropy Loss	Accuracy	CV Epochs
1	64/32/0.001/64	0.70	0.27	0.36	0.88	0.434	87.48	21
2	128/32/0.001/64	0.74	0.28	0.44	0.86	0.435	87.36	11
3	256/32/0.001/64	0.60	0.26	0.37	0.59	0.432	87.46	9
4	128/64/0.001/128	0.75	0.26	0.42	0.96	0.432	87.54	13
5	128/32/0.001/128	0.71	0.42	0.48	0.98	0.433	87.44	11
6	128/32/0.0005/128	0.64	0.28	0.43	1.27	0.434	87.48	12
7	64/32/0.0005/128	0.60	0.30	0.38	0.62	0.434	87.39	33
8	64/32/0.0005/64	0.62	0.34	0.44	0.82	0.434	87.43	33
Standard errors (N=2,000)		0.40	0.19	0.24	0.33	-	-	-

889

890 *Table 5: Hyperparameter configurations and model metrics for the attention based imputation model. Metrics were calculated*
891 *based on a sample size of N=2,000 diary days. Furthermore, a mean standard error due to the sample size of 2,000 diary days*
892 *is given.*

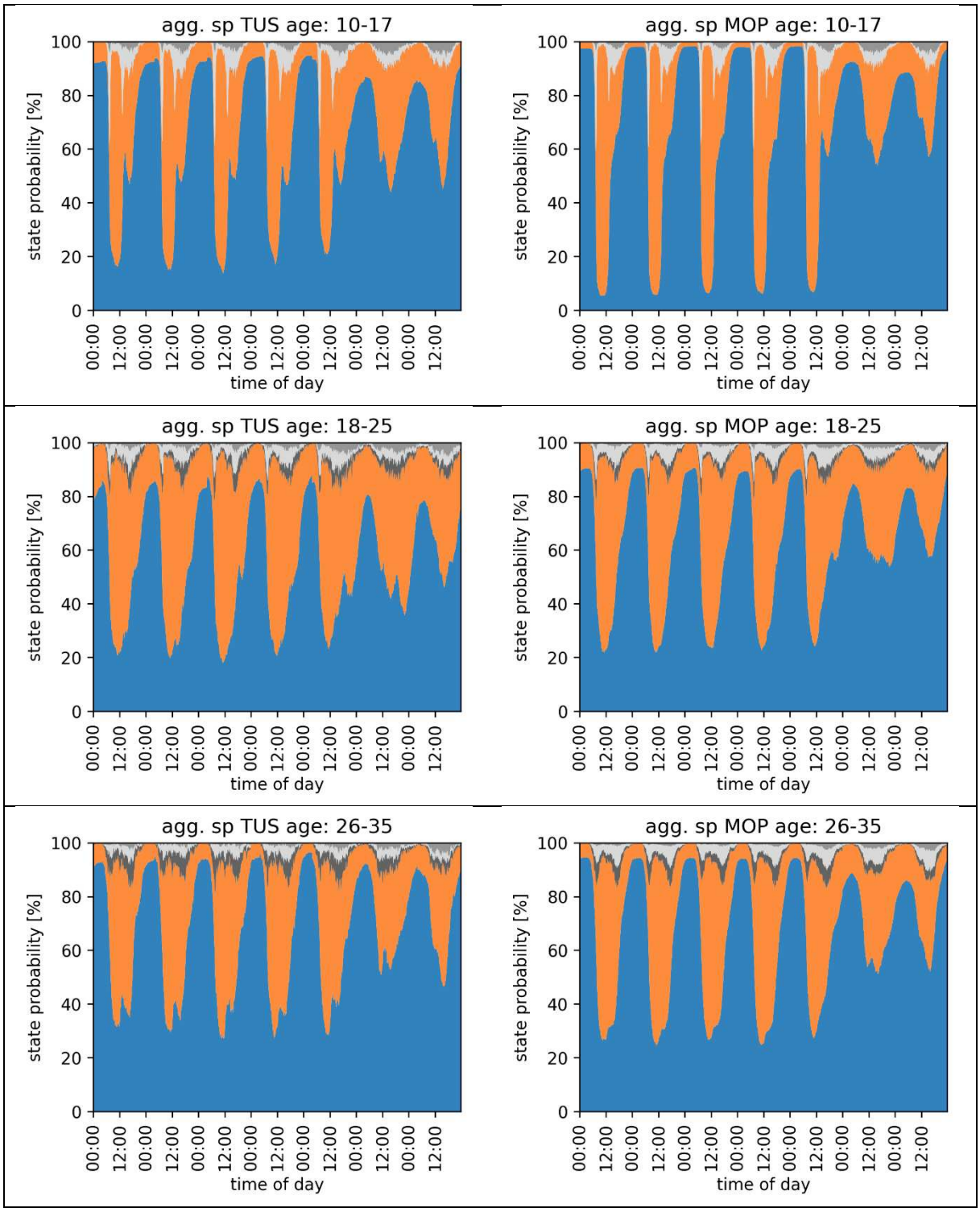
No.	Transformer layers/ D_model/ Learning rate/ Batch size	Sp rmse	Sd rmse	Ac rmse	Na mae	Cross-entropy Loss	Accuracy	CV Epochs
1	1/64/0.001/256	0.58	0.39	0.50	0.50	0.469	86.97	158
2	4/64/0.001/256	0.58	0.39	0.44	0.62	0.436	87.32	22
3	4/64/0.001/64	0.57	0.38	0.36	0.90	0.436	87.35	8
4	4/64/0.001/128	0.63	0.39	0.46	1.05	0.438	87.31	12
5	4/64/0.0005/64	0.59	0.36	0.39	0.72	0.435	87.35	10
6	**4/64/0.0005/128**	**0.49**	**0.39**	**0.39**	**0.66**	**0.431**	**87.41**	**37**
7	4/14/0.0005/64	1.27	0.54	0.72	1.42	0.458	87.14	46
8	4/14/0.0005/128	0.84	0.60	0.61	0.65	0.459	87.14	47
Standard errors (N=2,000)		0.40	0.19	0.24	0.33	-	-	-

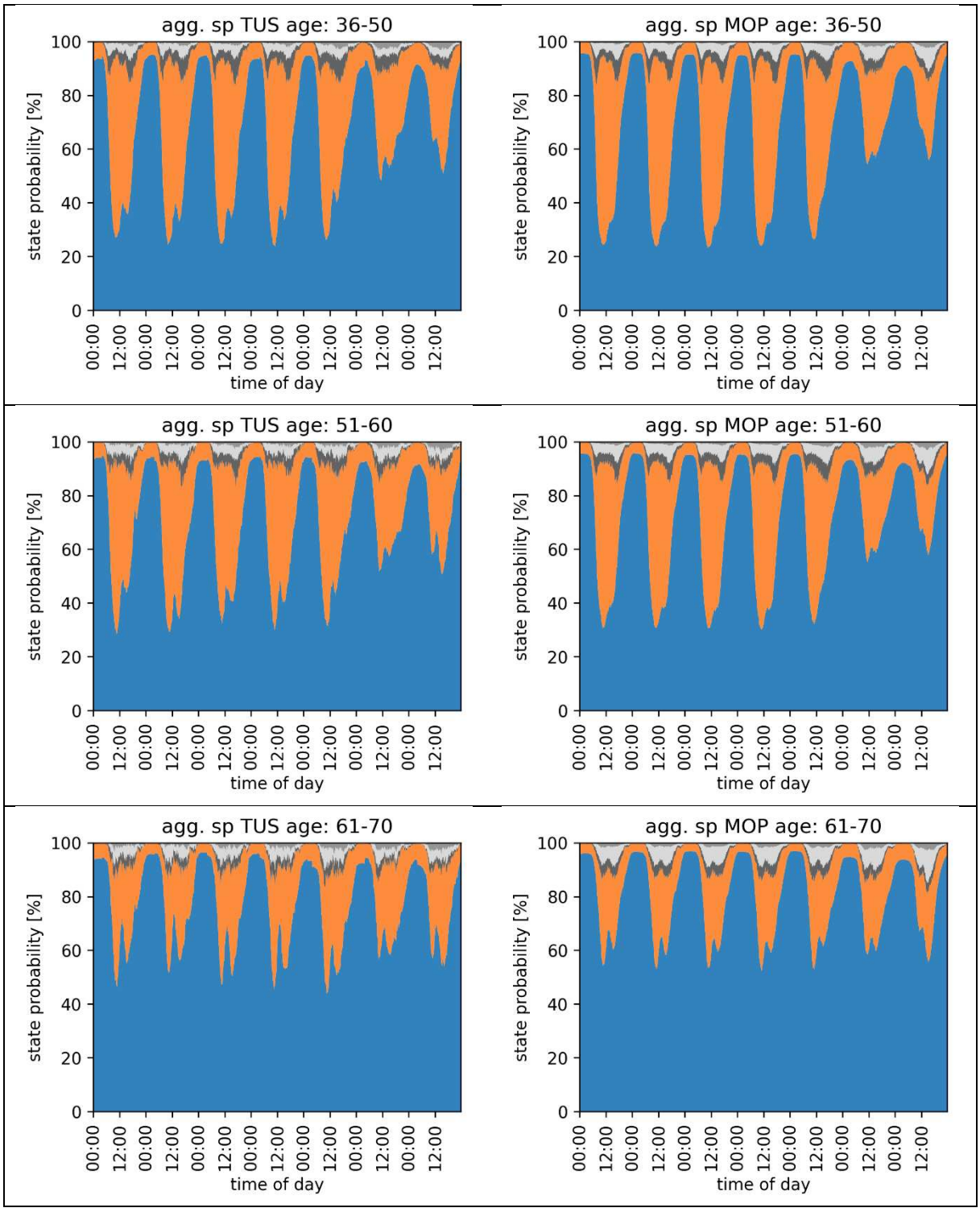
893

894 *Table 6: Comparative presentation of the socio-demographic composition of the MOP and TUD data sets. The calculated rmse*
895 *of the aggregated state probabilities are calculated on the basis of the five aggregated states (home, outside, mobile (car*
896 *driver), mobile (co driver), mobile (rest)). For the calculation of the rmse between the synthetic profiles and the MOP and TUD*
897 *data, synthetic data with the same socio-demographic characteristics as in the comparison data sets were generated.*

Age	**<18**	**<26**	**<36**	**<51**	**<61**	**<71**	**>=71**
Samples MOP	1971 (7.4%)	1430 (5.4%)	2288 (8.6%)	6107 (22.9%)	5132 (19.3%)	5809 (21.8%)	3873 (14.6%)
Samples TUD	2169 (18.2%)	1106 (9.3%)	1140 (9.6%)	4080 (34.2%)	1654 (13.9%)	1167 (9.8%)	494 (4.1%)
rmse sp MOP/TUD	4.0%	3.8%	2.3%	1.9%	2.2%	2.7%	2.8%
rmse sp syn./MOP	1.7%	1.6%	1.3%	0.9%	1.1%	0.7%	0.9%
rmse sp syn./TUD	3.9%	4.2%	2.1%	1.7%	1.9%	2.6%	2.7%
Job	**-**	**Full time**	**Part time**	**Students**	**Training**	**No job**	**Pensioner**
Samples MOP	212 (0.8%)	8853 (33.3%)	3627 (13.6%)	2759 (10.4%)	489 (1.8%)	2052 (7.7%)	8618 (32.4%)
Samples TUD	-	3938 (33.0%)	2599 (21.8%)	2214 (18.6%)	375 (3.1%)	1184 (9.9%)	1611 (13.5%)
rmse sp MOP/TUD	-	2.2%	2.5%	2.9%	3.8%	2.4%	2.4%
rmse sp syn./MOP	2.8%	1.1%	1.1%	1.3%	4.0%	1.0%	0.7%
rmse sp syn./TUD	-	2.1%	2.0%	3.6%	4.69%	2.1%	2.1%
rmse sp MOP/TUD (entire sample)				2.9%			
rmse sp syn./TUD (entire sample)				1.8%			
rmse sp syn./MOP (entire sample)				0.7%			

898 *Table 7: Comparative representation of the aggregated state probabilities of the TUD and MOP data sets for population*
899 *groups with different ages.*




agg. sp TUS age: 36-50
agg. sp MOP age: 36-50
agg. sp TUS age: 51-60
agg. sp MOP age: 51-60
agg. sp TUS age: 61-70
agg. sp MOP age: 61-70
state probability [%]
time of day
0
20
40
60
80
100
00:00
12:00

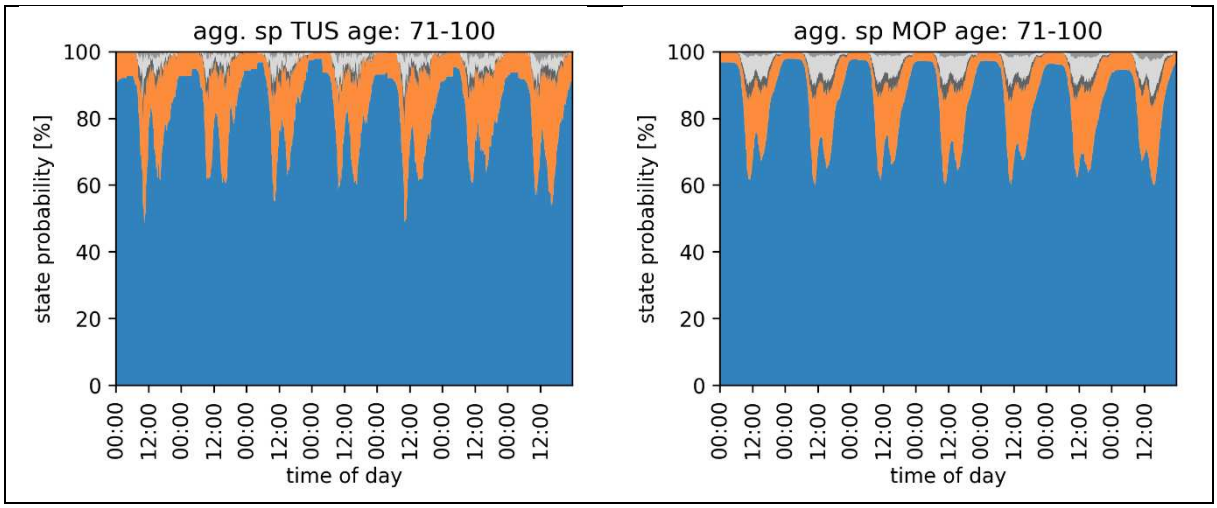
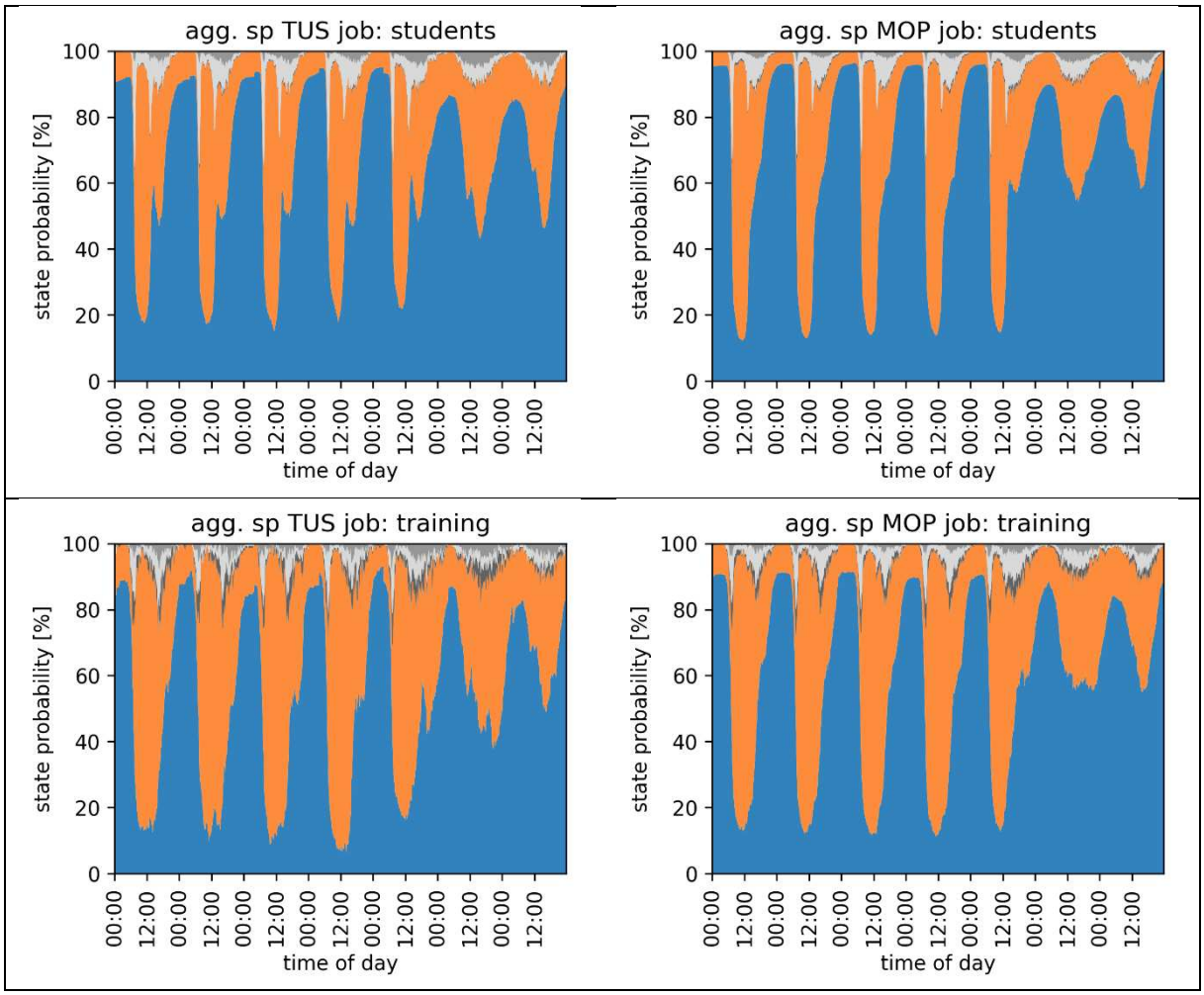
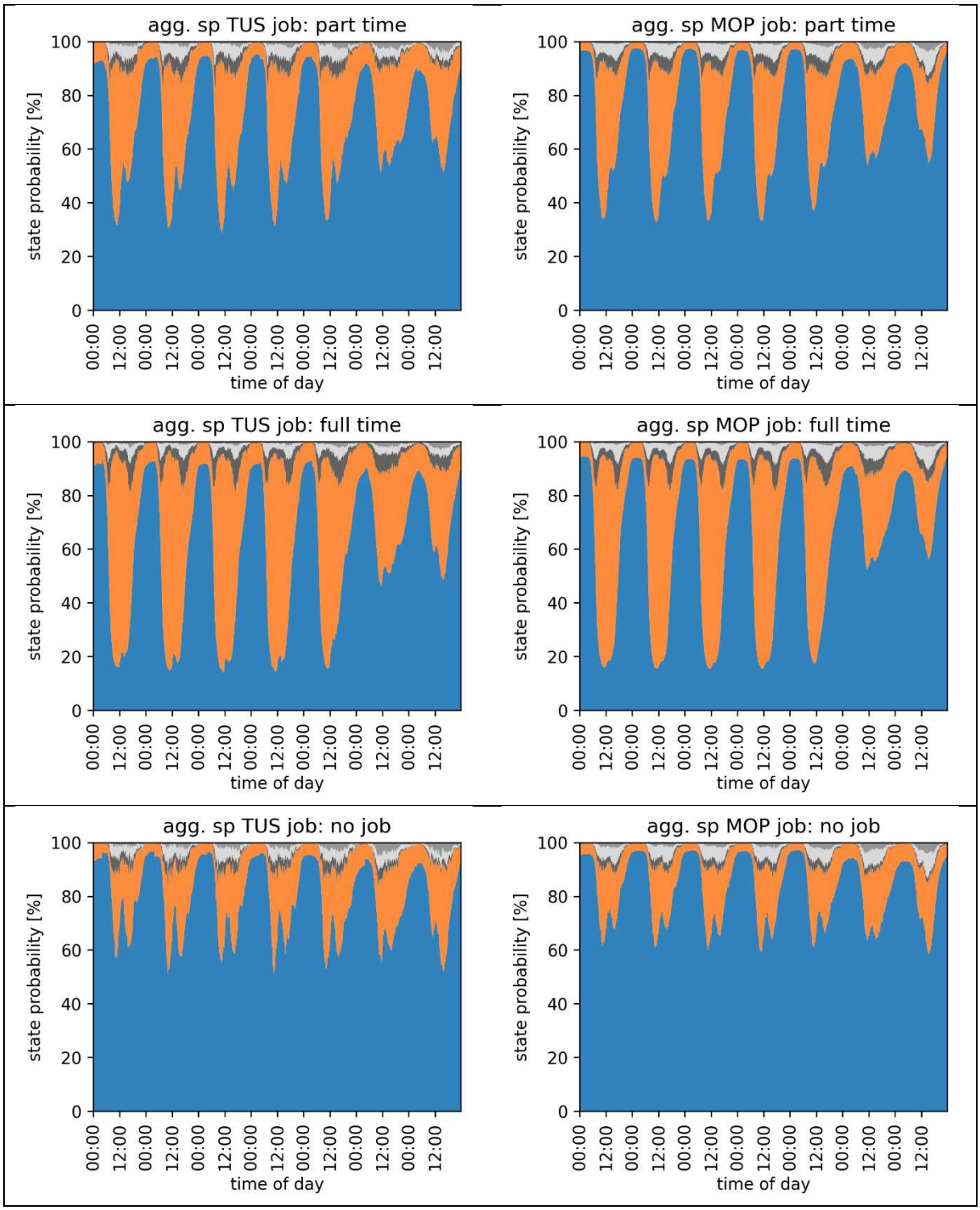
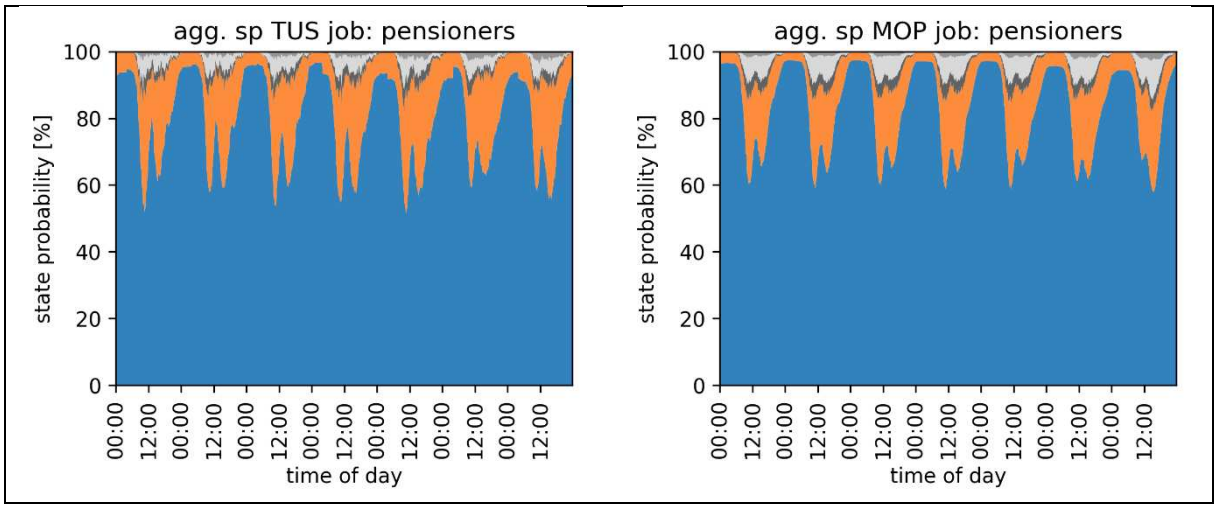


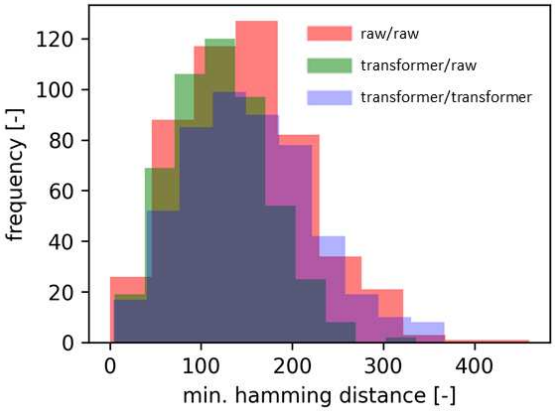
Table 8: Comparative representation of the aggregated state probabilities of the TUD and MOP data sets for population groups with different occupations.




agg. sp TUS job: part time
agg. sp MOP job: part time
agg. sp TUS job: full time
agg. sp MOP job: full time
agg. sp TUS job: no job
agg. sp MOP job: no job
state probability [%]
time of day
0
20
40
60
80
100
00:00
12:00



903



904

905 *Figure 19: Distribution of the minimum Hamming distances of the samples from dataset a (sample size N = 500) to the*
906 *samples in dataset b (dataset a/ dataset b)*

907

	Comments	Authors' response	Changes made, if any
	Reviewer #1		
1	The authors demonstrated the developed model are able to capture long-term dependencies in mobility and activity patterns at the same time adequately predict the diversity in behavior across the entire population. The capability is useful to model energy demand of buildings and mobility realistically. I believe that the developed methodology contributes significantly to the literature of the energy demand modelling. Therefore, I recommend the paper to be published as it is, with rechecking small errors. At least, the last paragraph of the introduction contains the error in referencing.	Thank you for your nice comment.	We checked the entire manuscript again for errors and removed the referencing error in the introduction (line 103).
	Reviewer #2		
2	This work is mainly about bridging the gap between social practice theory, energy related activity modelling and novel machine learning approaches. The recent machine learning method, namely, LSTM, is investigated in this work. The autoregressive model is employed, which generates synthetic weekly mobility schedules of individual occupants and thereby captures long-term dependencies in mobility behavior. The framework is clear. The reviewer would like to raise several comments.	Thank you for your comments. In addition to the LSTM-based approaches mentioned, attention-based neural networks are also presented. In addition to the autoregressive approach, an imputation model is presented, which enables the two data sets (TUD and MOP) to be linked.	No changes.
3	As the most successful machine learning methods, neural networks-based	Thank you for your comments. We totally agree that neural network-based deep learning models have the	In Section 3.1.3. we added the

	deep learning models have been widely studied across different fields. These methods generally follow an end-to-end framework. Namely, these methods have adaptively feature learning ability, regardless of data-preprocessing, manually feature extraction or selection.	ability to adaptively learn features and therefore manually feature extraction/selection is not needed while using these kinds of models. We see this ability as the greatest advantage of the model presented in this work compared to the regression and Markov models used so far in the literature in the field of activity modeling. However, since the data sets (Time Use Data and Mobility Data) used in this work are available in different formats, they must of course be brought to a uniform machine-readable data format before the data can be used for training these kind of models. Furthermore, for the connection of the data sets it was necessary to define an interSection of common activity states and meta-information on the basis of which the data sets can be combined with one another. To achieve this, the selected activity states and the selected meta-information were used. In summary, it can be stated that the data preprocessing was carried out with the aim of connecting the data sets using the two models presented. If we had "only" had the goal of reproducing a data set, then we totally agree with you, we would have had to provide the data to the neural network as unprocessed as possible. Regarding the additional sinusoidal and categorical positional encodings we provided to the networks: this information is necessary for the attention-based neural networks to learn the sequential structure of the underlying problem. This approach is used in most of the papers in literature which try to model sequential data with attention-based neural networks.	following sentences: *"In general, neural networks based machine learning methods have good adaptive feature learning ability. But in the present study the employed datasets are of a very different format, therefore they need to be aligned before the training."*
4	The reviewer thinks it is a significant improvement, compared with other machine learning methods.	Thank you for this comment. We do state that the "autoregressive part" of the presented approach is a significant improvement to the existing state of the art in the field of activity modelling which is defined by the	No changes.

		presented Markov and regression models presented in table 2.	
5	In this work, there are no comparisons between your LSTM model and other popular machine learning models.	Thank you for this comment. We present a sequential two step approach in our work consisting of an autoregressive model and an imputation model. For the autoregressive model we present an LSTM based neural network and an attention based neural network and compare these models with a first order Markov chain approach. Due to the similarity of the underlying problem, the selection of the methods used in this paper was based on the models that define the state of the art in the field of natural language processing. These are currently attention-based transformer architectures. Before that, LSTM based neural networks were used as described in Section 2. Of course, other approaches could be used for comparison, such as GRUs, which like LSTMs also represent recurrent neural networks. Due to the already large scope of the work and the expected similar results for LSTMs, this was not done. Furthermore, at the end of Section 2.3 we explain why GANs are not used.	In Section 3.2. we added the following sentences: *"Due to the similarity of the underlying problem, the selection of the methods used in this paper is based on the models that define the state of the art in the field of NLP. These are currently attention-based transformer architectures. Before that, LSTM based neural networks were used as described in Section* ***Error! Reference source not found.****."*
6	Besides, the existing comparisons seem to be unfair. The author illustrates the existing research Approaches in Table 1. What are the optimization measures of Markov Approach and which optimization strategy is adopted in this study? please add more details.	Thank you for this comment. We suppose that you consider the comparison of highly parametric neural networks with Markov models to be unfair. In this work, we propose the combination of two datasets together with new model approaches. By using both datasets we already have an advantage in information and by using neural networks which can capture the complex dependencies in the datasets we can generate synthetic data which combine the information of both datasets. Due to the specific nature of the overall approach presented, it is just possible to benchmark the single steps of the presented approach on their own against other approaches. At the end of Section 2.1, we discuss the general shortcomings of a Markov chain (exponential increasing number of	In Section 4.1 we added that it is a time-inhomogeneous Markov chain, since this information was missing until now: "*As a reference model for the presented autoregressive models, a* ***time-inhomogeneous*** *first order Markov model is used."*

		free parameters with order of model and the collection of all possible full high-order Markov chain models is limited and completely stratified). Due to these shortcomings, it is not possible to capture long-term dependencies over several days with Markov chains. As written in Section 4.1 (*As a reference model for the presented autoregressive models, a first order Markov model is used. The first order Markov model characteristics are representative for the models presented in Section 2.1, since marginal changes in the metrics can be achieved by using more complex Markov chains, but the basic problems remain (no long-term memory*)) we use a first order time inhomogeneous Markov model, which performs slightly worse than more complex Markov models, as can be seen from the papers discussed in Table 2.1, but the general shortcomings stay the same. Regarding the optimization measures of the Markov approach: as written above we used a first order time inhomogeneous Markov chain. Therefore, we calculated transition probabilities between all states for all time steps considered. Regarding the imputation approach we compare different bidirectional LSTM and attention based models, which seems to be appropriate, because these models define the state of the art (Cui et al. 2020; Chan et al. 2020; Richard et al. 2020; Ma et al. 2019; Sucholutsky et al. 2019).	
7	Although the model shows admirable performance, the Hyperparameter optimization in the study seems to be unreliable. What is the scope, basis and strategy of Hyperparameter adjustment? In particular, LSTM Units, Dense Grids and Transformer Layers.	Thank you for this comment. Due to the fact that our computational resources are limited, as stated in the beginning of Section 4, it was not possible to fully optimize all hyperparameters in a systematic procedure (grid search), as is the case in most machine learning based approaches. After experimenting with many different parameter settings (different activation functions,	We added the following paragraph to the Results Section: "*In addition to the learning rate and the batch size, the number of LSTM units was varied, which limits the complexity of the*

		dropout rates, embedding sizes,…) we came to the conclusion that the hyperparameters presented in Table 2 and Table 3 are the most relevant and interesting for the following reasons: - **LSTM units**: they define the memory capacity of the neural network. With a high number of LSTM units, it is possible for the network to build a more complex representation of the past in the internal state of the LSTM. Obviously, the computational complexity increases with more LSTM units. ➔ important for intertemporal dependencies - **Dense neurons of dense layer (LSTM) D_model (attention)**: more neurons enable learning more complex time step specific state representations. ➔ Important for timestep specific dependencies - **Transformer layers**: the number of transformer layers define the depth of the neural network. Deeper layers of neural networks learn higher class features of human behavior. The performance increase with depth can be seen in Table 3 and Figure 11. ➔ Important to learn high level representations	*internal state of the LSTM and is therefore important to capture temporal dependencies in behavior. The number of dense neurons (LSTM) or the model dimension (transformer) was varied to ensure that state-specific information is appropriately represented. Furthermore, the depth of the neural networks was varied, as this enables the neural network to learn higher level representations in human behavior."*

Chan, William; Saharia, Chitwan; Hinton, Geoffrey; Norouzi, Mohammad; Jaitly, Navdeep (2020): Imputer: Sequence Modelling via Imputation and Dynamic Programming. Online verfügbar unter http://arxiv.org/pdf/2002.08926v2.

Cui, Zhiyong; Ke, Ruimin; Pu, Ziyuan; Wang, Yinhai (2020): Stacked Bidirectional and Unidirectional LSTM Recurrent Neural Network for Forecasting Network-wide Traffic State with Missing Values. Online verfügbar unter http://arxiv.org/pdf/2005.11627v1.

Ma, Jiawei; Shou, Zheng; Zareian, Alireza; Mansour, Hassan; Vetro, Anthony; Chang, Shih-Fu (2019): CDSA: Cross-Dimensional Self-Attention for Multivariate, Geo-tagged Time Series Imputation. Online verfügbar unter http://arxiv.org/pdf/1905.09904v2.

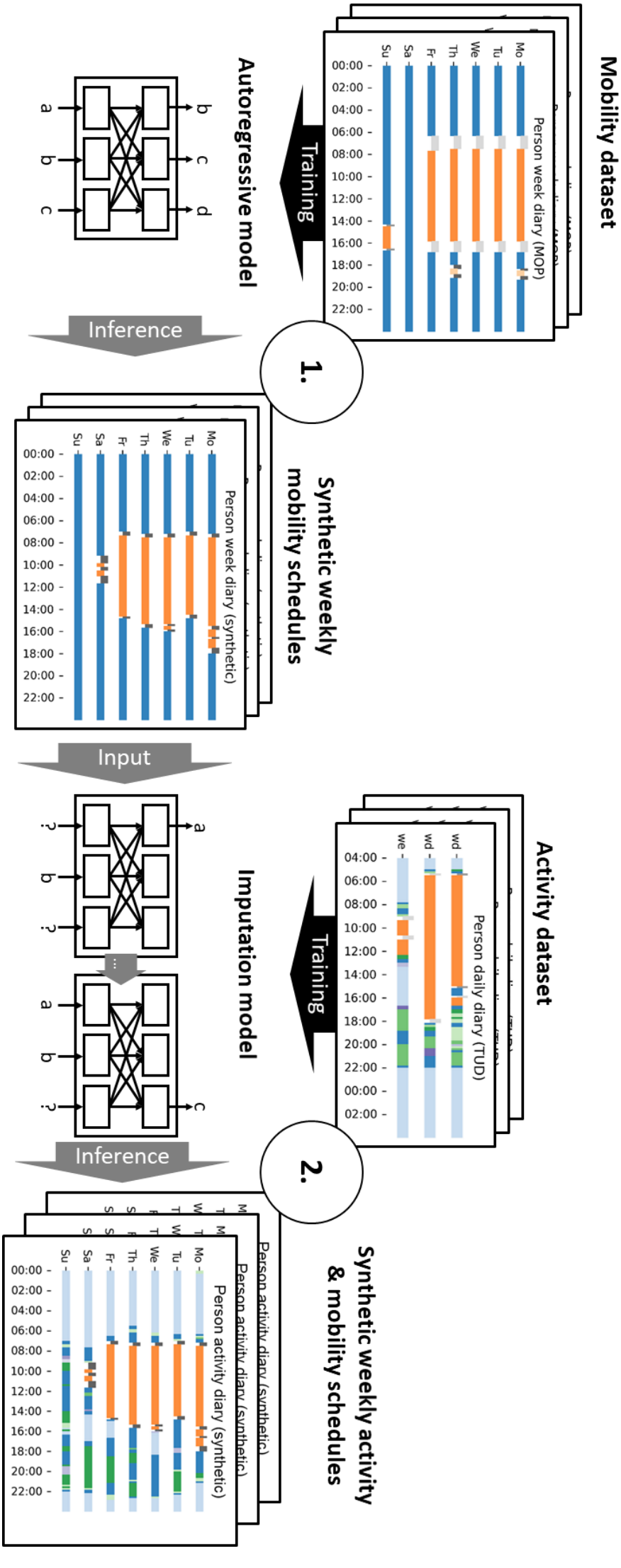
Richard, Wu; Aoqian, Zhang; Ihab, Ilyas; Theodoros, Rekatsinas (2020): ATTENTION-BASED LEARNING FOR MISSING DATA IMPUTATION IN HOLOCLEAN. In: *Proceedings of the 3rd MLSys Conference, Austin, TX, USA, 2020.*

Sucholutsky, Ilia; Narayan, Apurva; Schonlau, Matthias; Fischmeister, Sebastian (2019): Pay attention and you won't lose it: a deep learning approach to sequence imputation. In: *PeerJ Computer Science* 5 (1), e210. DOI: 10.7717/peerj-cs.210.

Highlights:

- Combining social practice theory, energy related activity modelling and novel machine learning approaches
- Neural networks can reproduce long-term dependencies in occupant behavior
- Advantages of longitudinal mobility data and detailed activity data are combined in one dataset
- Attention and long short-term memory based models are compared with Markov chains
- New data for consistent simulation of household electricity, heat and mobility demand

Graphical Abstract



Using neural networks to model long-term dependencies in occupancy behavior

Max Kleinebrahm[a], Jacopo Torriti[b], Russell McKenna[c,d], Armin Ardone[a], Wolf Fichtner[a]

[a] Chair of Energy Economics, Karlsruhe Institute for Technology, Hertzstraße 16, 76187 Karlsruhe, Germany

[b] School of the Built Environment, University of Reading, Whiteknights, PO Box 219, Reading RG6 6AY, United Kingdom

[c] DTU Management Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[d] School of Engineering, University of Aberdeen, Aberdeen AB24 3FX, United Kingdom

[a] Corresponding author: Max Kleinebrahm, max.kleinebrahm@kit.edu, +49 721 608 44691

Abstract

Models simulating household energy demand based on different occupant and household types and their behavioral patterns have received increasing attention over the last years due the need to better understand fundamental characteristics that shape the demand side. Most of the models described in the literature are based on Time Use Survey data and Markov chains. Due to the nature of the underlying data and the Markov property, it is not sufficiently possible to consider long-term dependencies over several days in occupant behavior. An accurate mapping of long-term dependencies in behavior is of increasing importance, e.g. for the determination of flexibility potentials of individual households urgently needed to compensate supply-side fluctuations of renewable based energy systems. The aim of this study is to bridge the gap between social practice theory, energy related activity modelling and novel machine learning approaches. The weaknesses of existing approaches are addressed by combining time use survey data with mobility data, which provide information about individual mobility behavior over periods of one week. In social practice theory, emphasis is placed on the sequencing and repetition of practices over time. This suggests that practices have a memory. Transformer models based on the attention mechanism and Long short-term memory (LSTM) based neural networks define the state of the art in the field of natural language processing (NLP) and are for the first time introduced in this paper for the generation of weekly activity profiles. In a first step an autoregressive model is presented, which generates synthetic weekly mobility schedules of individual occupants and thereby captures long-term dependencies in mobility behavior. In a second step, an imputation model enriches the weekly mobility schedules with detailed information about energy relevant *at home* activities. The weekly activity profiles build the basis for multiple use cases one of which is modelling consistent electricity, heat and mobility demand profiles of households. The approach developed provides the basis for making high-quality weekly activity data available to the general public without having to carry out complex application procedures.

Keywords: activity modelling, mobility behavior, neural networks, synthetic data

1. Introduction

In the course of the decarbonisation of domestic heat demand, it is expected that a large part of the heat will be generated by electricity (e.g. through heat pumps) (Paardekooper et al. 2018). In order to decarbonise the mobility sector, the aim is to increase the amount of electric vehicles in the European union from 1.3 million in 2020 to at least 33 million by 2030 (Transport & Environment 2020). Due to the expected developments, fundamental characteristics will change in the course of energy demand in the household sector. Furthermore, the introduction of stationary and mobile electricity storage systems as well as stationary heat storage systems enable the storage of energy over periods of single days and therefore open up flexibility potentials in the residential sector, which can support the integration of fluctuating renewable energies. To evaluate these flexibility potentials, fundamental relationships that shape household energy demand must be understood.

Occupant behavior has been identified as having a significant impact on household energy demand (Steemers and Yun 2009). Therefore, there has been an increasing research interest in the field of behavioral modelling over the last years with the aim to explain dynamics in residential energy demand based on energy related activities (Torriti 2014, 2017). A large number of studies focus on the modelling of activity sequences of single households or individuals with the objective to describe occupant behavior on an aggregated level for socio-demographic differentiated groups (Aerts et al. 2014; Flett and Kelly 2016; Richardson et al. 2008; Wilke 2013). Time use data (TUD) are used as a data basis, which provide information on the temporal course of occupant activities over single days and are available for various countries in the form of population representative samples (Eurostat 2000). Based on occupant behavior, different approaches were developed that connect occupant activities with electrical household appliances and thus generate synthetic electricity demand profiles (Yamaguchi et al. 2018). The aim of these studies is to gain a deeper understanding of household electricity demand in order to e.g. be able to evaluate device-specific efficiency measures, time-dependent electricity tariffs or load shift potentials.

However, TUD only provide information on activity patterns of individual days, therefore longer-term dependencies in mobility behavior and energy relevant *at home* activities that extend over several days are not captured in existing TUD based models. Figure 1 compares the autocorrelation of power consumption data generated on the basis of TUD with measured power consumption data. The autocorrelation in the generated data is underestimated. Especially, dependencies between subsequent days (48 lags) are not properly reproduced by the examined models.

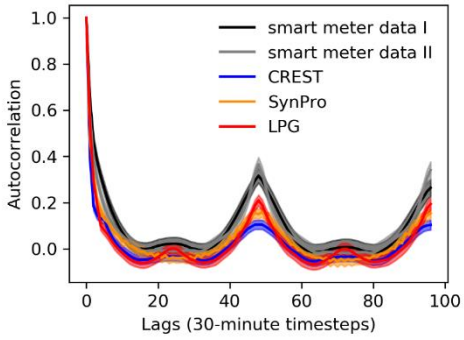


Figure 1: Mean autocorrelation and 95% confidence interval of electricity consumption profiles of the three load profile generators (LPG (Pflugradt 2016), CREST (Richardson et al. 2010), SynPro (Fischer et al. 2015)) and empirical smart meter data (I: HTW (Tjaden et al. 2015), II: (described in (Kaschub 2017))

Models based on device-specific power consumption data available over periods longer than one day are able to account for day-to-day variability in electricity demand (Yilmaz et al. 2017). However, due to the data underpinning these approaches, not much is known about the occupants and their

behavior, therefore it is not (easily) possible to calculate consistent heat and mobility demand profiles matching the electricity demand. One possible way to infer the occupancy behavior would be to use non-intrusive occupancy monitoring methods in order to calculate internal heat gains (metabolic gains and device-specific heat losses) (Chen et al. 2018). However, integrating demand through electrical vehicles would be another challenge.

The objective of this study is to develop a methodology that enables the generation of synthetic weekly activity schedules in which long-term dependencies in mobility behavior and energy relevant *at home* activities are captured on an individual level. These schedules can be used as a basis for generating consistent energy service demand profiles, taking into account heating, mobility and device specific energy service demand. In order to identify trends and potentials at the individual household level, like flexible charging behavior of electric vehicles, day-to-day variability in mobility patterns needs to be captured in the proposed approach. Therefore, novel machine learning based algorithms from the field of natural language processing (NLP) which are capable of capturing long-term dependencies in time series are transferred to the field of activity modelling. To answer the research question to what extent these algorithms are able to capture long-term dependencies in individual energy related occupancy patterns while maintaining the diversity of occupancy behavior on an individual and aggregated level, two behavioral data sets are combined in a two-step approach. Mobility data are used which provide information about weekly mobility patterns and combined with time use survey data which provide detailed information about daily activities (sleeping, cooking, eating, …).

The two-step approach enables to combine the advantages of mobility data (long-term dependencies in mobility behavior) with the advantages of TUD (detailed information about activities) and generates high quality weekly activity schedules. Novel machine learning algorithms which are used in the area of NLP are used for the first time to model occupancy behavior. These models have fundamental advantages over Markov chains, because they provide the capability to learn long term dependencies in time series. In comparison to existing approaches which were developed to reproduce aggregated occupancy behavior the proposed approach reproduces aggregated occupancy behavior and at the same time provides high quality individual activity schedules. Therefore, the synthetic activity schedules can be used to analyse trends in the household sector on an individual level and to examine their impact on an aggregated level at the same time. Due to the rich socio-demographic information in the underlying data sets, differences in behavior between socio-demographic groups can be analysed based on the synthetic activity schedules.

The paper is structured as follows. Section 2 presents an overview about current approaches to activity based residential demand modelling and gives a short introduction to the field of social practice theory. Furthermore, the latest developments in the field of NLP are summarized. Section 3 presents the mobility and activity data used in this work. Subsequently, two autoregressive models are presented for the generation of weekly mobility schedules and two imputation models are presented for enriching the synthetic mobility schedules with energy related activity information. The section concludes with a presentation of the metrics used to evaluate the activity plans. In Section 4 the generated activity schedules are evaluated. Finally, the results are discussed and an outlook on future work is given in Section 5 before conclusions are drawn in Section 6.

2. Introducing NLP to activity modelling

The majority of studies in the residential energy demand modelling literature simulate residential energy demand based on activity patterns. The most important data basis for modelling activity sequences is TUD. TUD are large-scale surveys which provide detailed information about how people spend their time. The mean of data collection is the time-diary instrument in which the respondents enter their activities in regular time steps. These so-called time-diaries contain activity sequences for

the period of usually one single day. When selecting households for the study, care is taken to select a sample of households representative of the population. Time diaries are collected for all persons in the households except for young children for usually one weekday and one weekend day to capture the differences between the days. Since TUD are collected in a harmonised procedure in most countries in Europe, these data provide a good basis for a variety of similar models for modelling activity sequences. In the following, different model approaches are presented which generate activity sequences based on TUD and similar activity-based data sets. Furthermore, the weaknesses of the models reviewed in the literature is described and a short insight into social practice theory is given. Finally, the field of NLP is briefly introduced due to similarities in modelling human behavior and language.

2.1. Markov chain based approaches

One of the most commonly used approaches to map activity sequences is to describe them as Markov chains. A Markov chain is a stochastic process that describes a sequence of possible states in which the probability of each state depends only on the previous states. The state space of a Markov chain describes the set of possible states and their corresponding state transition probabilities. The abstract idea behind the modelling of activity sequences that describe the behavior of individuals is that individuals go about their lives by transitioning between different elements of a set of potential states of activity (Ramírez-Mendiola et al. 2019). Richardson et al. have developed an occupancy model which uses a first order Markov chain and distinguishes between the states 'active at home' and 'not active at home' for each person of a household (Richardson et al. 2008). Based on aggregated household states they calculate transition probabilities in order to model the activity level of the household over the timeframe of one day. By modelling households in an aggregated way instead of individual persons, inter personal relations are better represented than in models where individuals are modelled individually (McKenna et al. 2015). First order Markov models are adequately suited to describe processes that fullfill the Markov property. The term Markov property refers to the memorylessness of a stochastic process. For a first order Markov model, this means that the transition to a subsequent state depends only on the current state and is independent of previously observed states in the evolution of the process. It is obvious that residential activity schedules represent more complex processes and therefore cannot easily be represented by a first order Markov model. To overcome this problem, a variety of more complex Markov models have been presented in recent years. In contrast to first order Markov models, so-called semi-Markov models determine not only the subsequent state but also the duration of the subsequent state. As this kind of models represent an improvement to first order Markov Chains, due to a better mapping of state durations, they are used in various studies for activity modelling (Aerts et al. 2014; Wilke 2013; Bottaccioli et al. 2019). Flett et al. (Flett and Kelly 2016) present a Markov model for occupancy simulation that uses transition probabilities which are calculated based on the current state and the length of the current state. By considering the state length of the current state, this model represents an improvement over previous models, so that this model cannot be called memoryless. The logical next step would be to develop higher order Markov models, which allow any number of past states to be taken into account when choosing the subsequent state. However, two serious issues can be associated with higher-order Markov chains. On the one hand the number of free parameters in the model increases exponentially with the order of the model and on the other hand the collection of all possible full high-order Markov chain models is limited and completely stratified (Ramírez-Mendiola et al. 2019). Ramírez-Mendiola et al. (Ramírez-Mendiola et al. 2019) addressed this issues by presenting a Markov chain model with variable memory length which allows the order of the model to vary during the evolution of the stochastic process. In order to find relevant portions of the past based on the influence on the outcomes of the transition probabilities to subsequent states the authors present a novel algorithm based on the Kullback-Leibler divergence and the log-likelihood test.

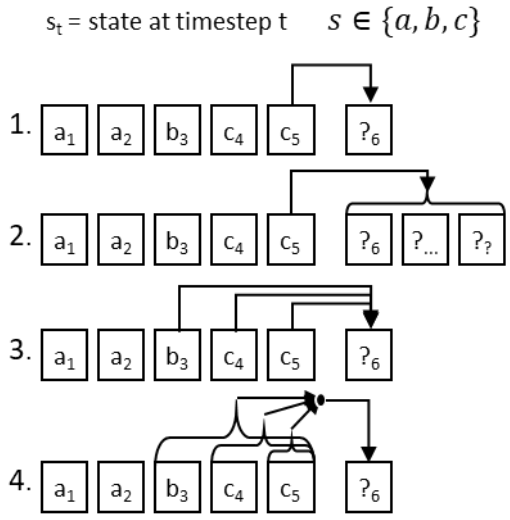


Figure 2: Graphical representation of the process of sequence generation with different kinds of Markov chains (1. first order Markov chain, 2. semi Markov chain, 3. higher order Markov chain, 4. Markov chain with variable memory length)

A graphical overview of the different Markov chain variations can be seen in Figure 2. It can be concluded that over the last few years more and more complex models based on Markov chains have been developed, which partly overcome the memorylessness problem. However, due to their structure, Markov models are only able to capture the states of the short-term past in order to predict subsequent states. Long-term relationships in daily schedules cannot be adequately represented by these types of models.

2.2. Timing of social practices

Markov chain approaches are based on the assumption that activities develop over time and are only dependent on the evolution of previous states. However, social practice theory literature points out that in order to understand people's daily/weekly schedules these should be treated as a whole (Shove et al. 2012; Torriti 2017). While practice theoretical accounts of social life vary, they remain consistent on at least two counts: (1) that practices are shared (socially/as part of the social i.e. performed by more than one person) and, because of that, (2) are repeated (performed more than once). If we also add that practices are connected and depend more and less on each other in being reproduced, it follows that we need to know more about how practices are repeated and with what effect for the relative strengths of their dependencies, connections, and extended relationships. In order to do justice to this statement in the patterning of activities, models must be developed which not only make it possible to capture connections between activities from the short-term past in order to predict the future, but also capture higher-level patterns which shape patterns of people's activities. In other words, models need to understand how temporal dynamics are embedded in the social world in order to understand how activities and thus energy consumption change and vary over time (Walker 2014). The majority of people structure their lives in daily rhythms, which are based on regular working hours, meal times and other constraints. These constraints form the basis for a certain degree of synchronization of social activities and thus for demand patterns (Walker 2014). Future models should be able to recognize and reproduce logical sequences in activity patterns, so that dependencies in activities are taken into account. For instance, food should first be prepared and then eaten.

Hilgert et al. (Hilgert et al. 2017) use a utility-based stepwise regression approach to generate weekly activity schedules for travel demand models. Due to the observation period of one week and the associated extended requirements for the mapping of activity sequences (day to day stability and variability of personal behavior), this approach differs from the approaches presented so far. Compared to Markov chain based approaches, activity sequences do not evolve over time but are the

result of regression based utility functions and time budgets. Based on Bowman (Bowman 1998) the construction process of activity schedules is split into smaller decisions due to the high complexity of constructing the entire schedule directly. These small decisions are then integrated downward vertically in the form of many logistic regression models. Due to the large number of regression models and their integration, many assumptions must be made when creating such a model, which increase the assumption bias. Future approaches should be less assumption driven to be easily transferable to different applications and datasets. To capture the high complexity of a complete activity plan without many intermediate steps, as described by Bowman, data-driven approaches could be used that need less assumptions and can capture complex relationships due to their structure. Table 1 gives an overview of the approaches presented in this section and compares them with the approach presented in this study.

Table 1: An overview of selected models for modelling occupancy behavior

Study	Database	Approach	Object of consideration	Country
(Richardson et al. 2008)	TUD	Markov - 1st order	Household	UK
(Wilke 2013)	TUD	Markov - semi	Individual	FR
(Bottaccioli et al. 2019)	TUD	Markov - semi	Individual	IT
(Aerts et al. 2014)	TUD	Markov - semi	Individual	BE
(Flett and Kelly 2016)	TUD	Markov - higher order	Individuals	UK
(Ramírez-Mendiola et al. 2019)	TUD	Markov - variable length	Individual	UK
(Hilgert et al. 2017)	MOP	Regression	Individual	DE
This study	**MOP + TUD**	**Neural networks**	**Individual**	**DE**

2.3. A brief review of natural language processing

The term natural language processing covers applications such as text classification, text understanding, text generation and text translation. NLP algorithms give machines the ability to read, understand and derive meaning from human languages. Over the last years NLP evolved from the era of punch cards and batch processing, in which the procession of a sentence could take up to 7 minutes, to the era of Transformer based model architectures like Googles BERT or OpenAIs GPT-3 with models up to 175 Billion parameters which are trained on large web corpora like Wikipedia and are able to generate articles which human evaluators have difficulty distinguishing from articles written by humans (Young et al. 2017; Brown et al. 2020; Devlin et al. 2018).

The first neural language model was based on a feed-forward neural network (Bengio et al. 2003). Vector representations of the n previous words are taken from a table and used as input in order to predict the probabilities of the following words. Nowadays dense vector representations of words or word embeddings are trained in an efficient way while training the neural network and are capable of capturing the context of words in a document (Mikolov et al. 2013).

From 2013 on neural network models in the form of recurrent neural networks (RNN), convolutional neural networks (CNN), and recursive neural networks got adopted in the field of NLP (Sutskever 2013; Kalchbrenner et al. 2014). RNNs are the obvious choice to deal with dynamic word sequences as they process the sequences from left-to-right or right-to-left and provide some kind of memory in the form of the hidden state (Elman 1990). RNNs in the form of long-short term memory networks (LSTM) proved to be more resilient to the vanishing gradient problem and therefore be able to better represent long-term dependencies in time series (Hochreiter and Schmidhuber 1997). The in 2014 presented sequence-to-sequence approach builds the basis for multiple machine translation applications. First, an LSTM-based encoder is used to compress an input sequence into a vector representation and then a decoder network, also based on LSTMs, predicts the target sequence step by step (Sutskever et al. 2014). The main shortcoming of the sequence-to-sequence approach is that

the input sequence needs to be compressed into a fixed-size vector. The Attention mechanism tackles this shortcoming by allowing the decoder to look back at the input sequence hidden states, which are provided as additional input to the decoder (Bahdanau et al. 2015). A rare feature of the Attention mechanism is, that it provides superficial insides about the learning process by providing information, through the attention weights, about which parts of the input are relevant for particular parts of the output. In 2016 Google presented their neural machine translation system which consisted of a deep LSTM network combining multiple encoder and decoder layers using residual connections and the attention mechanism (Wu et al. 2016). However, in 2017 the paper "Attention is all you need" was presented, which builds the basis for numerous transformer architectures which work on the principle of self-attention and define the state of the art in multiple NLP tasks (Vaswani et al. 2017; Brown et al. 2020). It was shown that the sequential nature can be captured by only using attention mechanisms and positional encodings without the use of RNNs. Due to the fundamental constraint of sequential computation of RNNs, it is not possible to parallelize training, therefore it is hard to learn on long sequences. Transformer models are fully based on fully connected layers and can be easily parallelized. Since 2017 multiple different transformer based architectures were introduced, consisting of multiple encoder and/or decoder blocks and an increasing number of trainable parameters (Wolf et al. 2020). In figure 3 the model architecture of a sequence to sequence RNN based model is compared to the model structure of an attention based transformer, consisting of an encoder and decoder block.

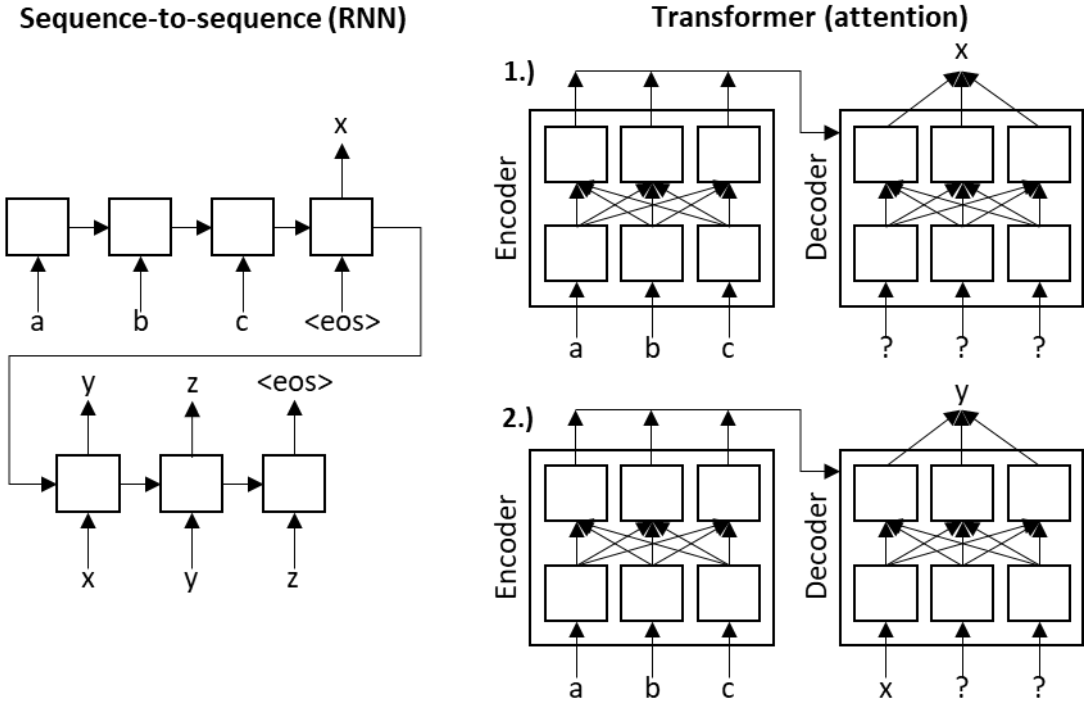


Figure 3: Abstract graphical representation of the RNN based sequence-to-sequence architecture (left) (Sutskever et al. 2014) and an encoder/decoder based transformer architecture on the right (Vaswani et al. 2017)

Adversarial learning methods have gained increased intention especially in the area of image processing/generation and have also been used in different forms in NLP over the last years. Generative adversarial networks (GANs) for example are able to generate synthetic data with similar statistical properties as real data by using two neural networks, a generator and a discriminator (Goodfellow et al. 2014). The generator produces synthetic data and the discriminator classifies generated data as fake and real data as real. Both networks are trained in an iterative way while trying to minimizes the reverse Kullback-Leibler divergence. Therefore, in comparison to the previously presented model architectures, GANs are not trained by maximum likelihood estimation (MLE) and thus are said to be less vulnerable to suffer from the exposure bias in the inference stage: the model generates a sequence iteratively and predicts next token conditioned on its previously predicted ones

that may be never seen in the training data (Yu et al. 2017). With that in mind many GAN based architectures were developed for natural language generation based on the approach presented in (Yu et al. 2017) which combines GANs with a reinforcement learning policy in order to deal with the differentiability problem. However, it was shown that MLE based approaches still dominate GANs when quality and diversity metrics are taken into account (Caccia et al. 2020). Therefore, GAN architectures are not considered further in this work, even if they form a promising basis for future work.

3. Data and Methodology

The German Mobility Panel (MOP) and German Time Use Data are used as an exemplary data source for analysing activity patterns in this study. In Section 3.1 the data preparation of the two data sets is described and the processed data is visualized. Further on, Section 3.2 presents the methodology developed to generate weekly activity schedules. Finally, Section 3.3 describes the metrics that are used to evaluate the activity plans.

3.1. Data

3.1.1. German mobility panel

The MOP collects information on the mobility behavior of the German population every year since 1994. About 1,500 to 3,100 persons (10 years and older), who make up about 900 to 1,900 households, fill out travel diaries over a period of one week. The travel diaries contain information about all trips during the week (start and arrival time, distance, modes used, purpose). In addition, socio-demographic information and information on refuelling behavior are recorded in the form of personal, household and fuel diaries. The survey is conducted every year in autumn to avoid distortions caused by holidays. The data is representative of the travel behavior of the German population. The Institute for Transport Studies at the Karlsruhe Institute of Technology is responsible for the implementation and design of the survey (Weiß et al. 2016; Zumkeller, Chlond 2009). Due to changes in the survey design, data from the surveys from 2001 to 2017 are used in this study.

3.1.2. German time use survey

For the analysis of energy relevant activities, the German part of the Harmonized European Time Use Survey, supplied by the German Statistic Office, was used (Destatis 2006; Eurostat 2000). Since the current version of 12/13 incorrectly recorded the location of the people, this data is not used. The data set contains activity diaries and socio-demographic information for 11,921 individual persons (age > 10 years) out of 5,443 households. Most of the participants provided diaries on two weekdays and one weekend-day in a 10-minute resolution. In this study time dependent data about primary activities and location as well as socio-demographic information are used.

3.1.3. Data preparation

In general, neural networks based machine learning methods have good adaptive feature learning ability. But in the present study the employed datasets are of a very different format, therefore they need to be aligned before the training. In order to create activity plans from the travel diaries, the basic dataset consisting of 833,986 travel entries for 35,014 person-weeks is converted into weekly activity plans with a time resolution of 10 minutes. The generation of activity plans is inspired by Hilgert et al. (Hilgert et al. 2017). In a first step, person weeks with missing or unrealistic entries are eliminated so that finally 26,610 person-weeks can be used for further analysis. Based on the travel entries and their purpose, states are determined for each time interval of the week. The choice of the initial state is based on the final state of the time series. Subsequently, the data are aggregated from a 1-minute

resolution to 10-minute resolution, assuming the state that is most frequently taken in the respective 10-minute interval. The reason for the reduction of the temporal resolution of the data is, on the one hand, the increased information density, since machine learning algorithms have problems with sparse data. On the other hand, TUD data are also recorded in 10-minute resolution.

The diary entries in the German TUD consist of more than 200 activity codes describing activities in the everyday life of human beings. Before the diary data is used as input for further processing, these activities are aggregated to activities relevant for household energy demand. The choice of activities is based on similar studies (Fischer et al. 2015; Richardson et al. 2010). The aggregated activities are visualized in Figure 4. In the upper two figures, the time course of the aggregated state probabilities of the two data sets is provided over a week. The lower two partial figures show example artificial activity plans for individual persons. Interday dependencies in behavior from Monday to Friday can be easily recognized from the visualization of the mobility schedule. The example activity plan, on the other hand, provides detailed daily information on energy-related home, sleep and mobility activities. Further comparative analyses based on socio-demographic characteristics of the data sets can be found in Section 5 and in the appendix.

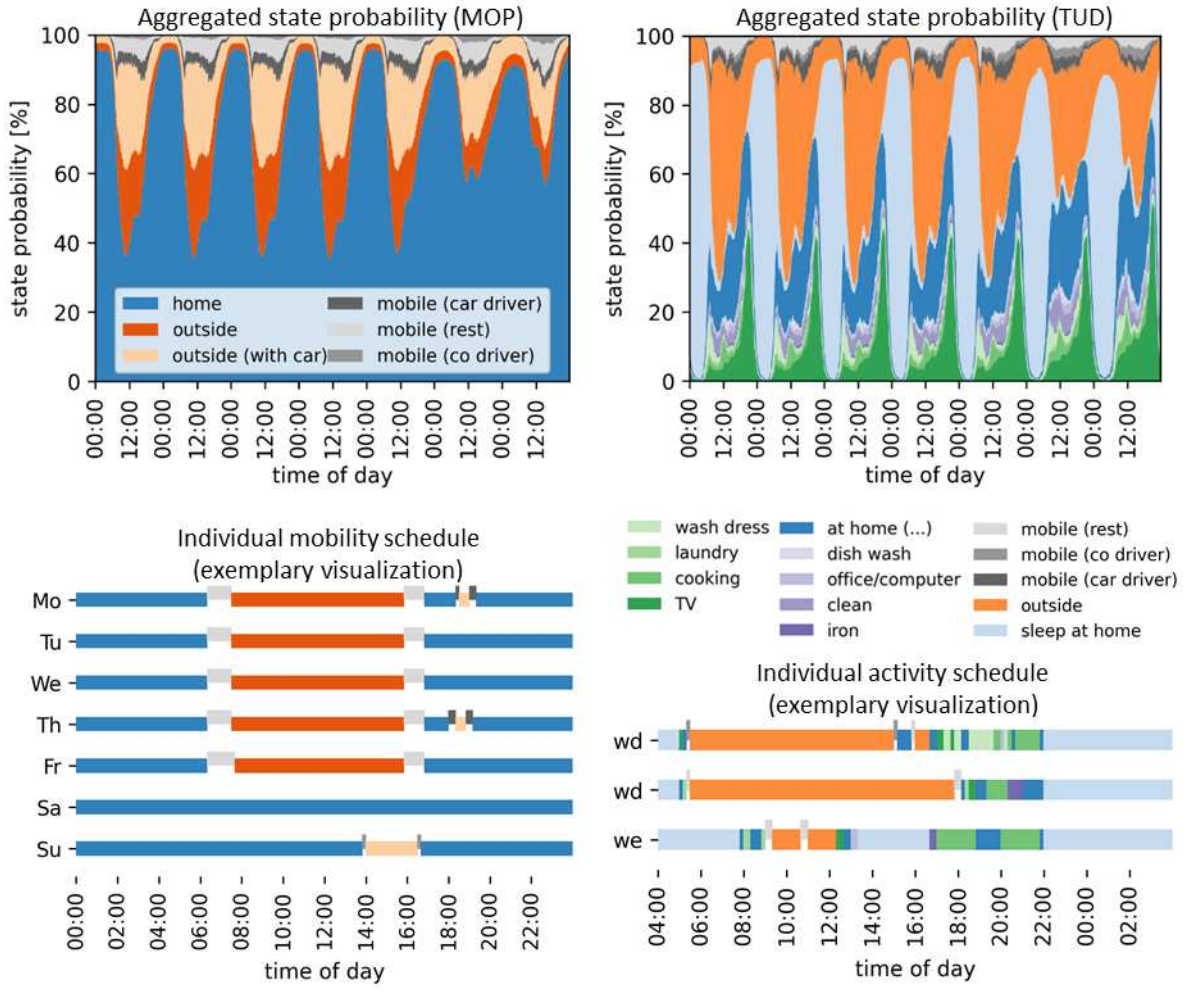


Figure 4: Visualization of aggregated state probabilities and exemplary artificial individual diary entries based on the MOP (Weiß et al. 2016) and the TUD (Destatis 2006)

3.2. Methodology

The approach for the generation of weekly activity schedules with a time resolution of 10 minutes is presented in Figure 5. In the first step, weekly mobility schedules of individual persons from the German Mobility Panel are used as input data. The objective of the first step is to generate synthetic

mobility schedules with statistical properties similar to the empirical schedules. The developed approaches are autoregressive. This means that it is assumed that the choice of the next mobility state ms_{t+1} only depends on all the states $ms_{0\ldots t}$ that have already been observed. In Section 3.2.1, an LSTM-based and an attention-based approach for sequence generation of mobility states are presented. Due to the similarity of the underlying problem, the selection of the methods used in this paper is based on the models that define the state of the art in the field of NLP. These are currently attention-based transformer architectures. Before that, LSTM based neural networks were used as described in Section 2.3.

The objective of the second model step is to enrich the synthetic mobility plans with energy-related *at home* activities. For this purpose, two imputation models are presented in Section 3.2.2. Bidirectional LSTM model architectures are compared with attention-based architectures. Time Use Survey data from individuals are used to train the models. During the prediction process, the synthetically generated weekly mobility schedules are fed into the imputation model as input and the *at home* state is replaced by energy-relevant activities. A graphic representation of the step by step procedure of the autoregressive and imputation models can be found in Figure 7 a.

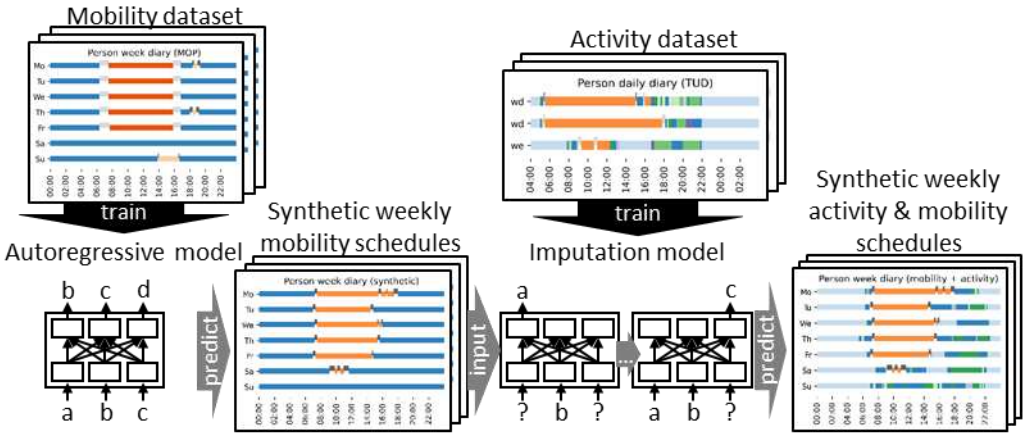


Figure 5: Two-step model approach for generating weekly activity schedules

3.2.1. Autoregressive models for weekly mobility schedule generation

To generate high-quality mobility plans on an individual level and at the same time representative mobility plans on an aggregated level that adequately describe the diversity of human behavior, approaches are required that capture the complex relationships in human behavior. In contrast to the Markov-based approaches used in the majority of the studies described in Section 2.1, LSTM and attention-based approaches can take into account longer-term time dependencies in the timing of individual activities due to their different memorisation mechanisms. While in Markov models probabilities are assigned to individual activity sequences and thus the number of free parameters increases exponentially with the order of the model, these kind of models are not suitable to take into account long-term dependencies in behavior between single days (Ramírez-Mendiola et al. 2019).

LSTM based models process time series sequentially and take as input the current state vector $x_t \in \mathbb{R}^d$ the hidden state vector $h_{t-1} \in \mathbb{R}^h$ and the cell state vector $c_{t-1} \in \mathbb{R}^h$. The dimension of the hidden state and the cell state vector h is the number of LSTM units which define the memory capacity of the LSTM cell. The cell states are adjusted every timestep using different gating mechanisms (input gate, output gate, forget gate) and activation functions. Due to the additive structure of the LSTM cells they partly solve the vanishing gradient problem and therefore are able to capture long-term dependencies in time series (Hochreiter and Schmidhuber 1997).

Attention based models do not process time series sequentially and therefore are suitable to better parallelize the learning process. The time dependencies between individual time steps are learned from scratch. To make this easier, positional encodings are added to the individual states in this study, which provide information about the relative position of the state in the time series. To calculate the masked dot product attention matrix, the matrices $Q, K, V \in \mathbb{R}^{T,d}$ (query, key, value) and the mask $M \in \mathbb{R}^{T,T}$ are required as input according to Figure 6. In the case of self-attention Q, K, V are the same. The mask shown in Figure 6 is a look ahead mask. The masked (black) cells contain high negative values and are added to the scaled result of the matrix multiplication of Q and K. The subsequent use of the softmax function prevents to put attention on dependencies between already observed and future states. The Softmax function transforms a T-dimensional vector with real components into a T-dimensional vector $\sigma(z)$ also as a vector of real components in the value range [0, 1], where the components add up to 1.

$$\sigma(z)_t = \frac{e^{z_t}}{\sum_{t=0}^{T} e^{z_t}} \quad t = 1, \dots, T \tag{1}$$

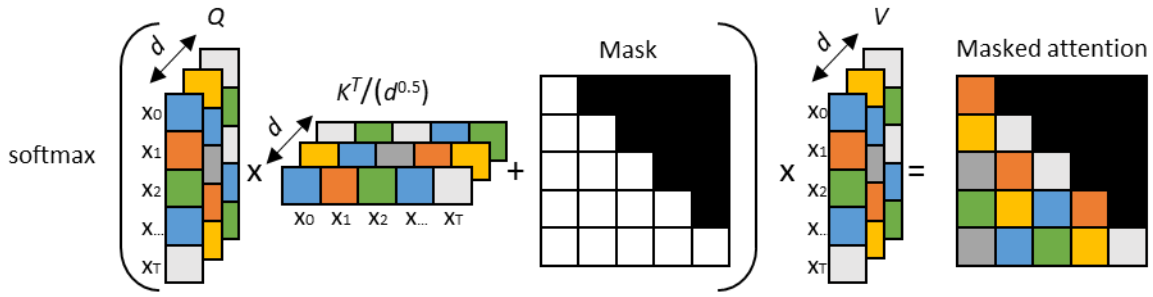


Figure 6: Illustration of the masked scaled dot product self-attention mechanism of an autoregressive model based on (Vaswani et al. 2017)

Before the dependencies between individual states can be learned in the LSTM/attention layers, layers must be introduced that use all the available information of a single state as input and learn its state representation in a multidimensional space.

Figure 7 b./c. show the different kinds of input provided to the autoregressive and imputation models and their first layers. Input to the autoregressive model is provided in the form of the mobility state $ms_{p,t}$, the time of the day/week τ_t, the day of the week d_t of person p at timestep t and as socio-demographic information $sd_{i,p}$. The time of the day/week is translated into a sinusoidal positional encoding using periods of one day/week. This is a typical approach to provide information about cyclical characteristics in time series (e.g. daily/weekly patterns) to the model. All other model inputs ($ms_{p,t}$, d_t, $sd_{i,p}$) are categorical and are therefore inserted into an embedding layer. Through the embedding layer the categorical information is mapped into a m-dimensional continuous space. The weights of the embedding layer and therefore the way the categorical variables are represented in the m-dimensional space are learned during the training process of the model. Further on, all the time step specific information are concatenated. The input time series is shifted one time step to the right ($t = 0 \dots T-1$) and starts with a dummy time step at $t = 0$, which is composed of a start token consisting of the start time and day and socio demographic information of the specific person. This training method is called teacher forcing (Williams and Zipser 1989).

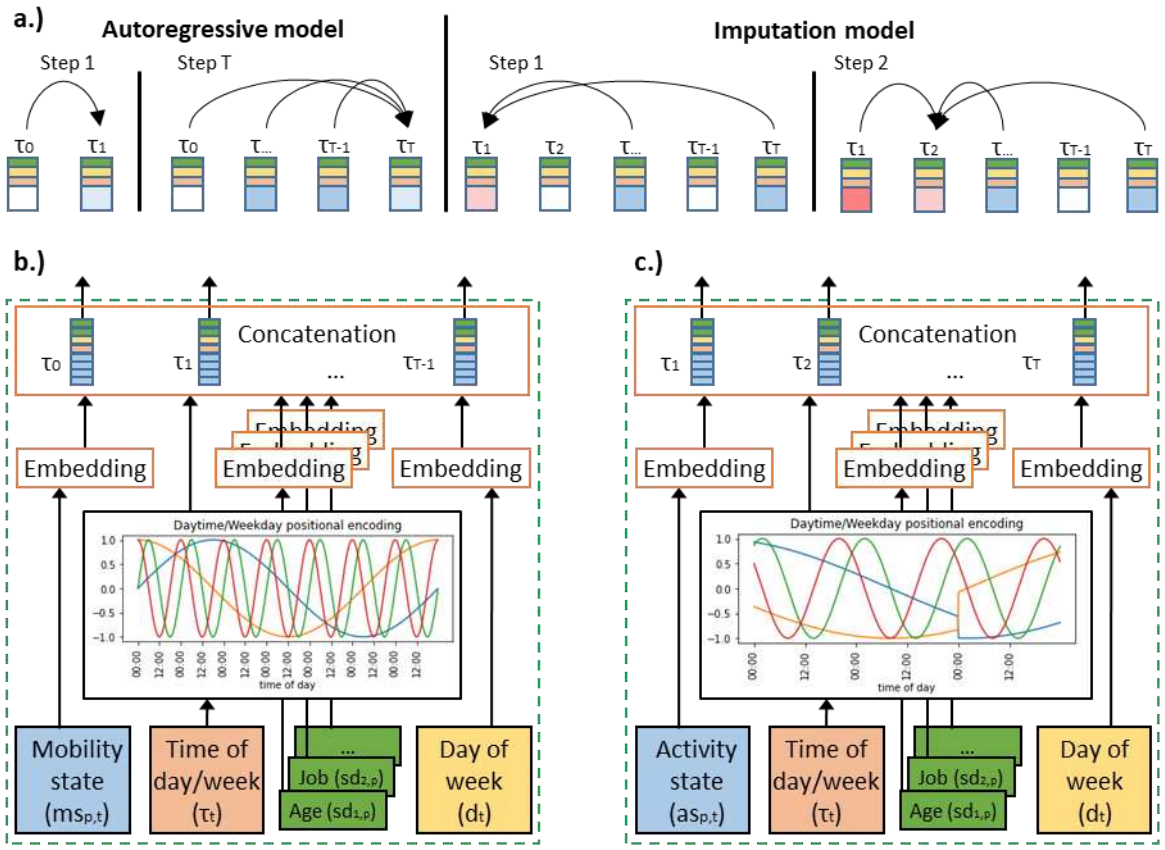


Figure 7: a.) Illustration of the relevant time step specific dependencies in the autoregressive and imputation models, b./c.) training input of the autoregressive/imputation (b./c.) models and visualization of their first layers

Figure 8 a. describes the central components of the LSTM based autoregressive model. After concatenating the time specific information, the vector state representations are fed into a linear dense layer before the state representations are inserted into a sequence to sequence LSTM layer. The final dense layer contains $|ms| = 6$ neurons which represent the probabilities (logits) of each mobility state $ms_{p,t}$ $(t = 1 \dots T)$.

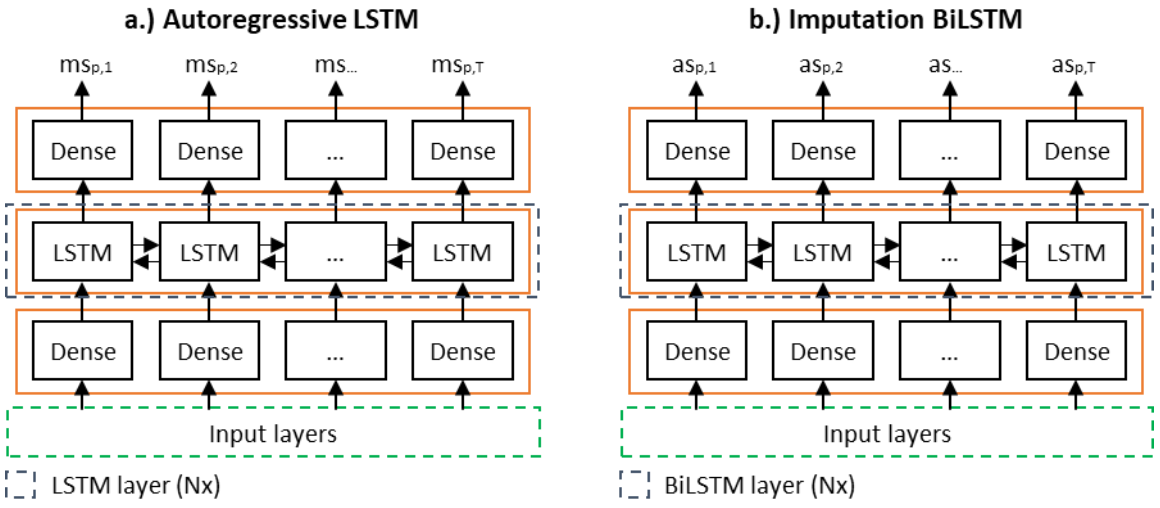


Figure 8: a.) LSTM based autoregressive model architecture and b.) BiLSTM based imputation model architecture

Figure 9 describes the architecture of the attention based transformer model. The transformer layer consists out of three linear dense layers for Q, K, V, the attention layer consisting of the scaled dot - product attention and two feed forward dense layers with dropout similar to (Vaswani et al. 2017).

Both models are trained by minimizing the cross entropy loss between the ground truth and the predicted probabilities.

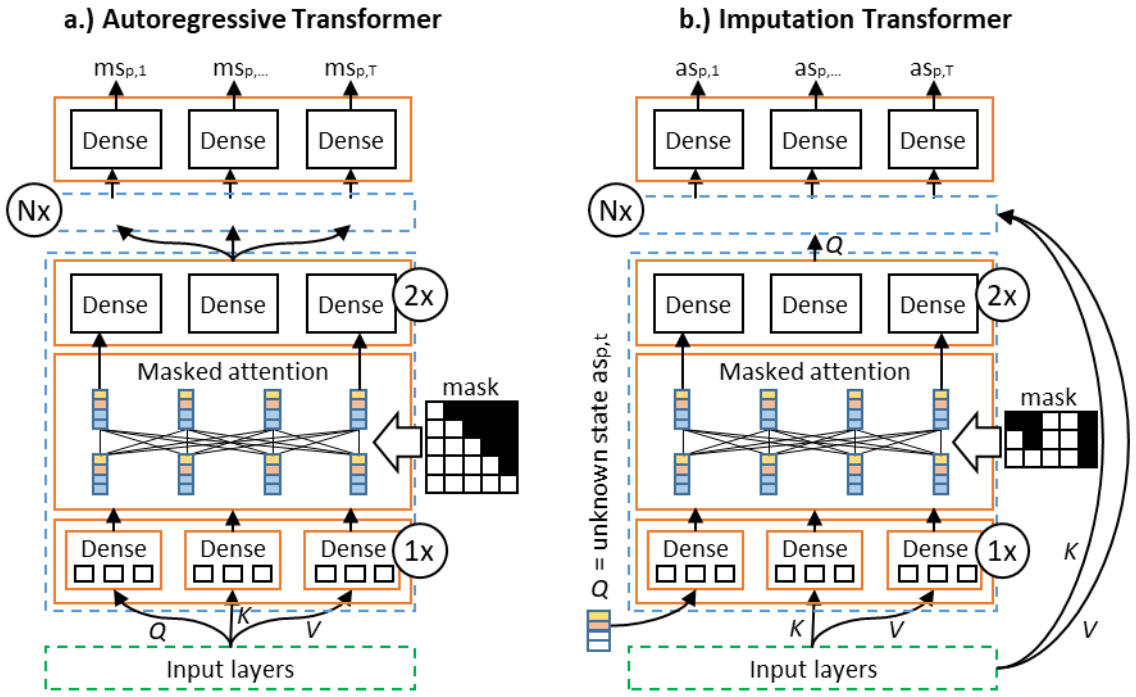


Figure 9: a.) Transformer based autoregressive model architecture and b.) Transformer based imputation model architecture (residual connections are not visualized)

3.2.2. Energy related activity imputation / enrichment

In the second model step, the generated weekly mobility plans are enriched with energy-related activities. A bidirectional LSTM model (Figure 8 b.) is compared with an attention-based transformer model (Figure 9 b.). In contrast to the first model step, information about individual mobility behavior over the entire week is already available when the first "at home" activity is estimated, this information has an impact on the activity choice. The procedure of the prediction process of the imputation model can be found in Figure 7 a.

As input data during the training process, the model is provided with activity time series of individual persons over 3 days (2x weekday, 1x weekend), the time and day of the week as well as socio-demographic parameters (job, age). The time step specific input processing can be seen in Figure 7 c. In contrast to the autoregressive models, the imputation models do not necessarily receive consecutive days as input, as this is not possible due to the structure of the time use survey. The connection between the three respective days is learned in the training process and applied to a whole week in the imputation process. In contrast to Figure 8 a., it can be seen in Figure 8 b. that the bidirectional LSTM architecture also takes future states into account when predicting the current state.

In contrast to the autoregressive transformer model, the imputation transformer does not use self-attention. The query vector Q of the first transformer layer contains the information about the unknown home states (unknown state, time, day, socio-demographic information). The key and value vector are identical and contain information about the mobility states of the three days (during training) or the week (during prediction). During the training process, *at home* activities of the TUD are masked and fed to the model as input. In all of the following Transformer layers, the output of the previous Transformer layer represents the query vector Q. The imputation models are trained using the cross entropy loss function.

3.3. Metrics

To evaluate the models presented, metrics must be introduced on the basis of which the model output can be assessed on an individual and aggregated level. The metrics presented below are generated and visualized at constant intervals during the training process.

The model-specific metrics are the cross entropy loss, which is minimized during the training process, and the model accuracy which provides information about how well the model predicts the next state. For the evaluation of the generated activity schedules, metrics are used to assess whether the proposed models reflect the variability in human behavior. Furthermore, metrics describing the variability of intrapersonal behavior are used to assess the consistency within a person's activity plan.

The aggregated state probability (sp) describes the aggregated probability $sp_{s,t}$ of a state $s \in S$ at time step $t \in T$ over a sample with the sample size N.

$$sp_{s,t} = \frac{\sum_{i=1}^{N} x_{i,s,t}}{N} \quad \forall\, s \in S, t \in T \tag{2}$$

State durations (sd) are calculated for all states $s \in S$ and are visualized by their cumulative distributions. The distribution of the duration of states can be used as a first indicator to evaluate the models with regard to the consideration of long-term time dependencies. For the evaluation of the intrapersonal variability within an activity schedule, the number of activities per week (na), the autocorrelation (ac) and the Hamming distance (hd) are calculated for each activity schedule of a sample. The autocorrelation is calculated for each activity state and each individual and is used to obtain information about the regularity of activities. The Hamming distance is calculated between all working days $d \in \{1 \dots 5\}$ of the week and thus provides information about the similarity of the daily behavior of individuals.

$$hd_n = \sum_{d_1=1}^{5} \sum_{d_2=1}^{5} \left|\left\{t \in \{1, \dots, T_d\} \middle|\, s_{d_1,t} \neq s_{d_2,t}\right\}\right| \quad \forall\, n \in N \tag{3}$$

From the variability of these metrics (na, ac, hd), information about the diversity in behavior can be obtained.

4. Results

The results presented below were calculated with an XLA compiler and a "Tesla V100-SXM2-16GB" GPU in Tensorflow 2.3. To provide the models from overfitting, the data sets are randomly split up into training data (9-fold cross validation → 80 % training, 10 % validation) and test data (10 %).

4.1. Mobility schedule generation

As a reference model for the presented autoregressive models, a time-inhomogeneous first order Markov model is used. The first order Markov model characteristics are representative for the models presented in Section 2.1, since marginal changes in the metrics can be achieved by using more complex Markov chains, but the basic problems remain (no long-term memory). The introduced metrics are visualized in Figure 10. All metrics shown are calculated based on a sample size of N = 2,000 unless explicitly stated otherwise. The course of the aggregated state probability of the state *outside* deviates only slightly from the empirical course. The averaged root mean square error (rmse) over all states of the aggregated state probability is 0.53 % and tends towards zero with increasing sample size. From the course of the cumulative state durations of the state *mobile (car driver)* and the other states shown in Figure 16 it can be observed that the state durations of the schedules produced by the first order Markov model partly deviate from the empirical data. Furthermore, the distribution of the Hamming distance and the autocorrelation clearly differ between the data generated by the Markov model and the empirical data, which is reflected in large deviations in the rmse of the autocorrelation and the

mean absolute error of the Hamming distance. The peak in the autocorrelation in mobility behavior after 144 lags (one day) describes daily mobility patterns in the mobility behavior of individual persons. This peak, which can be clearly identified in the empirical data, is not represented in the synthetic mobility schedules of the Markov model. Compared to the empirical distribution, the distribution of the Hamming distances is shifted to the right, towards higher distances. Consequently, subsequent days of single individuals differ more from one another than in the empirical data. The distribution of number of activities per week indicates that the Markov model matches the empirical data well on average, but the boxplot indicates that the diversity in behavior deviates from the one observed in the empirical data.

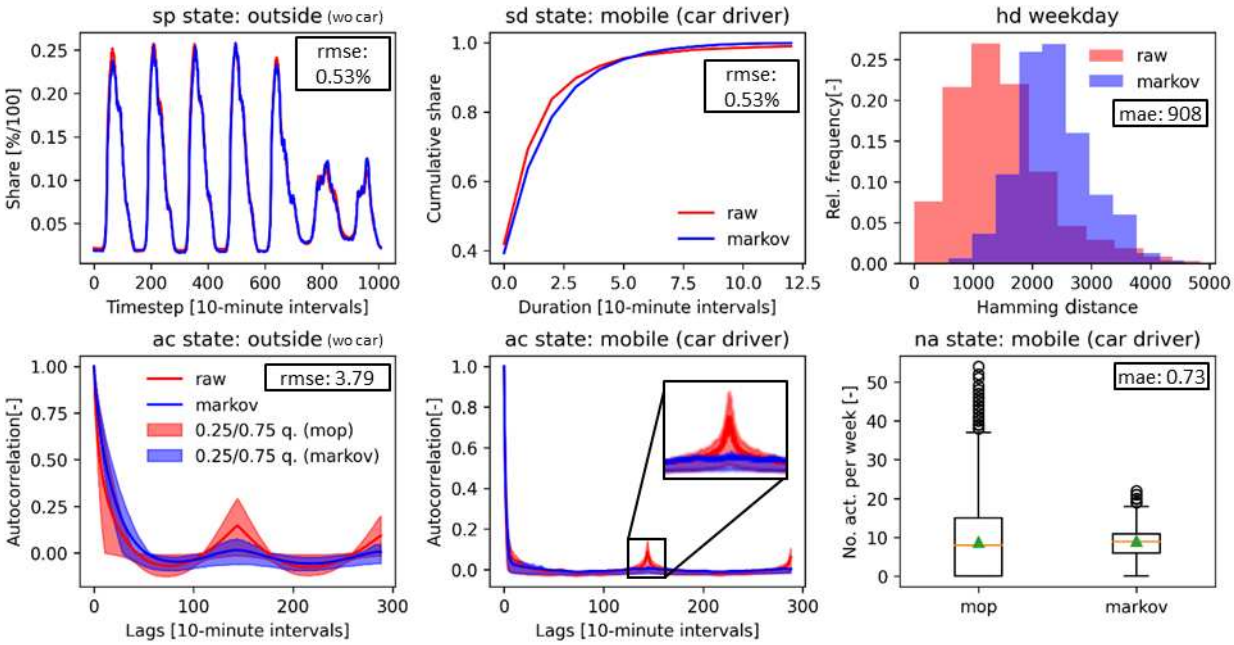


Figure 10: Visualization of the metrics for empirical MOP data (N = 26,610) and data generated with a first order Markov model (N = 2,000) (blue). The shown state dependent errors are calculated over all states and the mean is presented.

The autoregressive models presented in Section 3.2.1 are trained to predict the multinomial state distribution of the subsequent state. To achieve this, the cross entropy loss is minimized. Figure 11 and Figure 12 describe the course of the cross entropy loss during the training process. An epoch is defined as one training step of the nine-fold cross validation. After nine epochs, the training and validation data set are reshuffled and divided into nine new participations. During the training of the attention-based models, the loss function converges continuously for the training and test dataset. In the LSTM-based model, however, it can be seen that the course of the loss and accuracy function of the test data set diverges from the course of the training and validation data after around 14 epochs. From this point on, the model overfits on the training data and the training process can be stopped. In order not to use over-trained models, the weights of the model are saved at constant intervals during the training process. Furthermore, the development of the model accuracy during the training process is shown. This converges to a value of approx. 96.3%. This means that 96.3% of the time the correct value is predicted in the training process. Of course, the prediction is easier during the night when people are asleep than, for example, in the afternoon when there are many changes in activity. Figure 11 shows the course of the cross entropy loss for two model configurations, with one transformer layer and with four transformer layers. By increasing the depth of the neural network, the model can better map the complexity of mobility behavior. However, only marginal improvements can be achieved by further increasing the number of transformer layers from four to eight (Table 3). Since the performance of the models presented depends heavily on the choice of hyperparameters, various parameter settings were tested during the training phase for the LSTM and the attention based models. The parameter settings

varied during the training process and the corresponding metrics can be found in Table 2 and Table 3. In addition to the learning rate and the batch size, the number of LSTM units was varied, which limits the complexity of the internal state of the LSTM and is therefore important to capture temporal dependencies in behavior. The number of dense neurons (LSTM) or the model dimension (transformer) was varied to ensure that state-specific information is appropriately represented. Furthermore, the depth of the neural networks was varied, as this enables the neural network to learn higher level representations in human behavior. The results of the parameter variations show that the attention-based models are slightly superior to the LSTM-based models in most metrics, consequently, the attention-based model no. 3 from Table 3 is used for the presentation of the mobility schedule specific metrics.

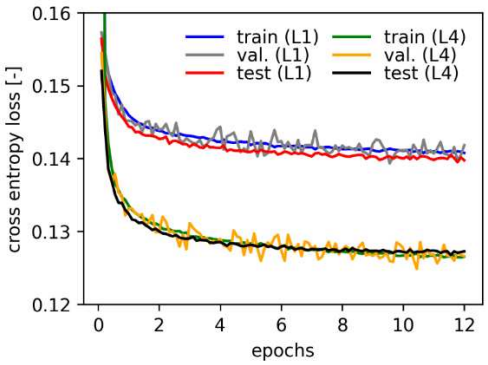


Figure 11: Loss development during training of the autoregressive transformer (L1/L4: 1/4 transformer layers)

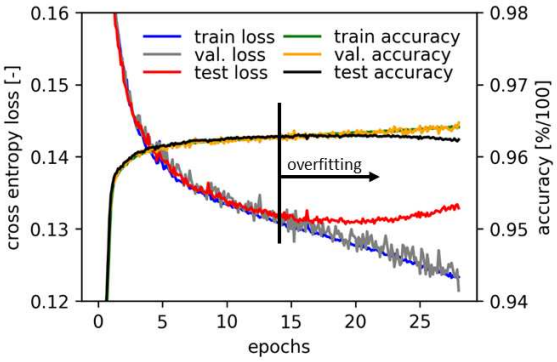


Figure 12: Loss and accuracy development during LSTM training

Selected mobility schedule specific metrics for the attention based autoregressive model described in Table 3 (model no. 3) are presented in Figure 13. A holistic overview of all metrics for all states can be found in the appendix (Figure 16). In contrast to the first order Markov model, the aggregated state probability is represented slightly worse by the attention based model. The rmse of the state probability averaged over all states and time steps is higher than the error of the first-order Markov model for all the models shown in Table 2 and Table 3 in the appendix. The Markov error corresponds to the standard error that arises with a sample size of 2,000. The standard error was calculated by randomly sampling 2,000 samples 30 times from the entire population and calculating their deviation from the metrics of the entire population (N = 26,610). The mean value of the error of the 30 samples is called the standard error. The mean absolute error of the number of weekly activities in the attention-based model is also higher than that of the Markov model (3.6 > 0.73). The diversity of the number of weekly activities is, however, recorded much more accurately by the attention-based model, which is shown in the lower right illustration in Figure 13 for the state *mobile (car driver)* and in Figure 16 for all other states. While the machine learning models presented in this work have slight deviations in the description of the averaged behavior and therefore perform slightly less accurately than Markov models, the mobility schedules generated differ fundamentally on the individual level, which is shown by the distribution of the cumulative state durations, the Hamming distance between weekdays and the autocorrelation of the individual states. Using the Hamming distance and the autocorrelation, it can be clearly seen that day-to-day dependencies in behavior are very accurately taken into account by the models presented in this work. In order to be able to adequately capture daily rhythms in mobility behavior, it is very important that the peak in the autocorrelation graph is captured well after 24 hours (144 10-minute time steps), which can be seen in the bottom center graph in Figure 13. From the course of the mean values and the ranges of the 25% / 75% quantile, it becomes clear that both these dependencies in the mean and in the spread are well represented across the entire population. These visual findings are also reflected in the significantly lower rmse of the autocorrelation compared to the Markov model (0.54 <3.79).

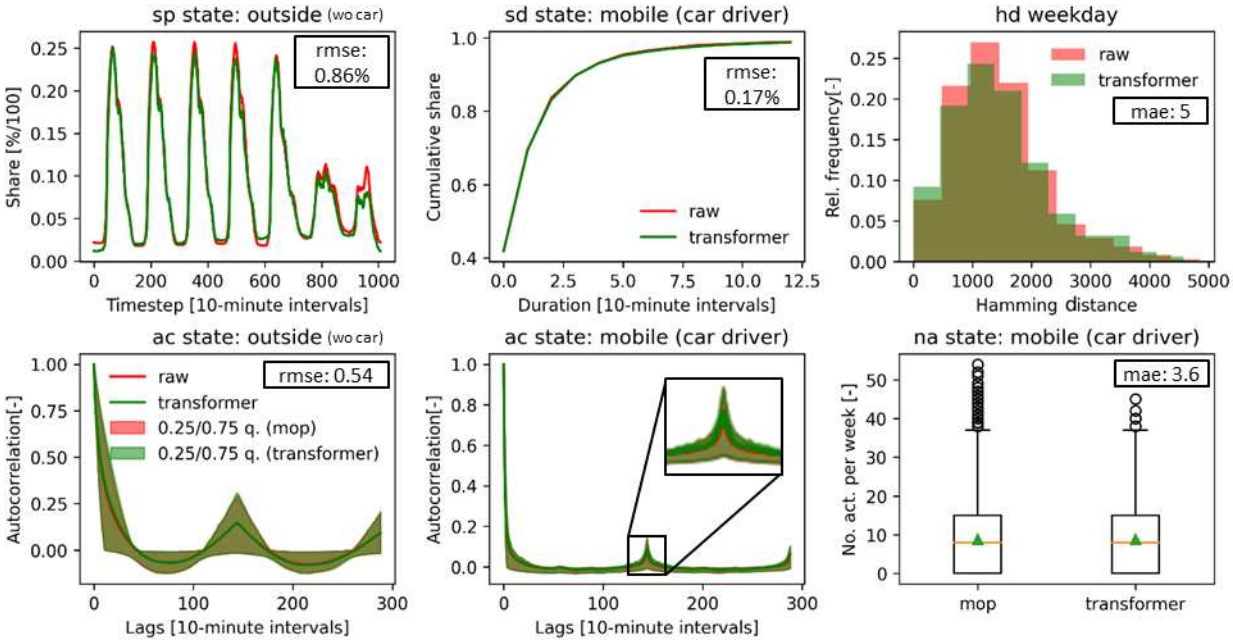


Figure 13: Visualization of the metrics for empirical MOP data (N = 26,610) and data generated with an attention based model (N = 2,000) (green). Model parameters can be seen in Table 3 (no. 3). The shown state dependent errors are calculated over all states and the mean is presented. The overlapping green and red ranges in the left-bottom and center-bottom graph describe the 25%/75% quantiles.

The difference between LSTM-based models and attention-based models is particularly evident from the autocorrelation peak in mobility behavior after 24 hours. LSTM models are also able to recognize relationships over such long periods of time, but in this work it was not possible to reproduce the peak as well with LSTM-based models as it can be seen in Figure 13 (bottom center) with the attention-based model. In addition to the low deviation of the mean error in the distribution of the Hamming distance (5 < 908), it can also be clearly recognized from the form of the distribution that the diversity in the profiles generated matches the real distributions much better than that of the Markov models, in which individual weekdays of a person do not have the similarities found in the empirical data.

4.2. Energy-related activity imputation

Since the model approach presented in this paper (step-by-step simulation of mobility behavior and subsequent enrichment of the results with energy-related activities based on different data sets) is new and no classical comparable applications in the field of behavioral modeling are known, only the results of the imputation models presented in Section 3.2.2 are benchmarked against each other in this section. As with the autoregressive models, the model performance of the imputation models is strongly dependent on the choice of hyperparameters. The parameters of the BiLSTM-based and the attention-based imputation model that were varied during the training process can be found in Table 4 and Table 5 in the appendix. To ensure that dependencies between time steps can be adequately captured by the model, sufficient amounts of LSTM units and attention layers must be provided. The dimension of the model must be chosen so that all time-step-specific information can be mapped well. In the following, the activity schedule-specific metrics for the attention-based model no. 6 from Table 5 are compared with the empirically collected TUD data. The metrics are visualized for specific states in Figure 14. A holistic overview of all metrics for all states and the development of the model loss and accuracy can be found in the appendix (Figure 17/Figure 18).

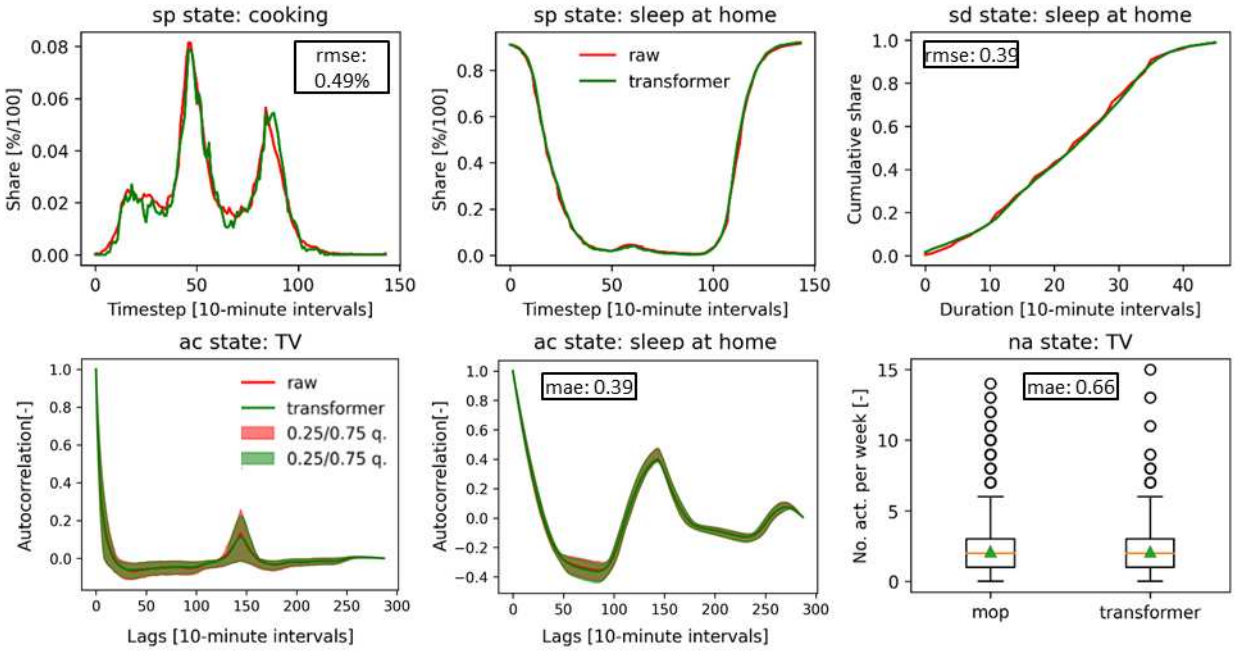


Figure 14: Visualization of the metrics for empirical TUD data (N = 35,691 dairy days) and data generated with an attention based model (N = 2,000 diary days) (green). Model parameters can be seen in Table 5 (model no. 6). The shown state dependent errors are calculated over all states and the mean is presented. The overlapping green and red ranges in the left-bottom and center-bottom graph describe the 25%/75% quantiles. The autocorrelation graphs were calculated based on the two work days over 288 10-minute timesteps.

Similar to the autoregressive models, it can be seen from the course and the rmse of the aggregated state probability that this differs slightly from the empirically collected data. The averaged errors over all states and time steps can be taken from Figure 14, Table 4 and Table 5 for the various model variants. The error in the simulation of the state durations, on the other hand, is smaller than that which occurs when modelling activities with a first-order Markov chain (no imputation model). Since the German TUD data set contains diary entries for three days of the week, the model can also learn day-to-day dependencies between energy-relevant activities. The autocorrelation graphs in Figure 18 show that the imputation model is able to recognize and reproduce these dependencies. For example, daily sleep rhythms can be reproduced in the synthetic data, which is another unique selling point of this work.

When comparing the metrics shown in Table 4 and Table 5, it is noticeable that the attention-based models perform slightly better in representing the aggregated state probability, while the BiLSTM-based models tend to map the duration of states and autocorrelation better. This could be attributed to the fact that when representing energy-relevant activities, short-term temporal dependencies between individual states are of higher importance than the one seen in the mobility schedules and the sequential character of the BiLSTM depicts these dependencies well, while attention-based models tend to capture individual states and their time-dependent probability of occurrence more strongly than short-term sequential dependencies.

4.3. Generation of weekly activity schedules

After the training processes of the autoregressive models and the imputation models have been described and evaluated in Sections 4.1 and 4.2, synthetic weekly activity plans are now generated for various socio-demographic groups and compared with empirical data. Table 6 in the appendix gives an overview of the socio-demographic composition of the empirical data. The age distribution of the MOP data shows that older population groups are overrepresented in contrast to the TUD data. Younger groups of the population such as students and part-time workers, on the other hand, are under-

represented. Due to the consideration of socio-demographic factors when coupling the data sets in the approach presented, a different distribution of the socio-demographic groups in the individual data sets is not problematic. When considering the sample sizes of the MOP and TUD data, it must be taken into account that the TUD samples, in contrast to the MOP samples, only consist of two to three days. The MOP data set with 10-minute time resolution has more than five times as many data points as the TUD data set. From the rmse of the aggregated state probabilities for the different socio-demographic groups, it can be seen that the data sets differ in some cases more strongly (rmse (age <18): 4.0%). In the synthetic profiles, the mobility behavior is generated on the basis of the MOP data, consequently, when looking at the rmse, fewer errors can be found between the synthetically generated data and the MOP data, both when looking at the socio-demographic groups in a differentiated manner and when looking at the aggregate as a whole dataset.

Finally, Figure 15 shows the course of the aggregated state probabilities over a week and two exemplary activity plans of synthetically generated schedules for two socio-demographic groups (age<18, full time employees). From the visualization of the aggregated state probabilities it can be seen that children under the age of 18 are mainly out of the home in the mornings and have two pronounced mobility peaks at around 8 am and 1 pm, while full-time employees are mainly outside during the day. Rhythmic behavior within the working days can be seen in the exemplary individual profiles. In the activity plan of the student on Friday morning, the student changes from an *at home* state to an *outside* state without a mobility activity in between. At first glance, this seems unrealistic, but these transitions can also be found in the empirical data due to the temporal aggregation of the mobility data over 10 minutes.

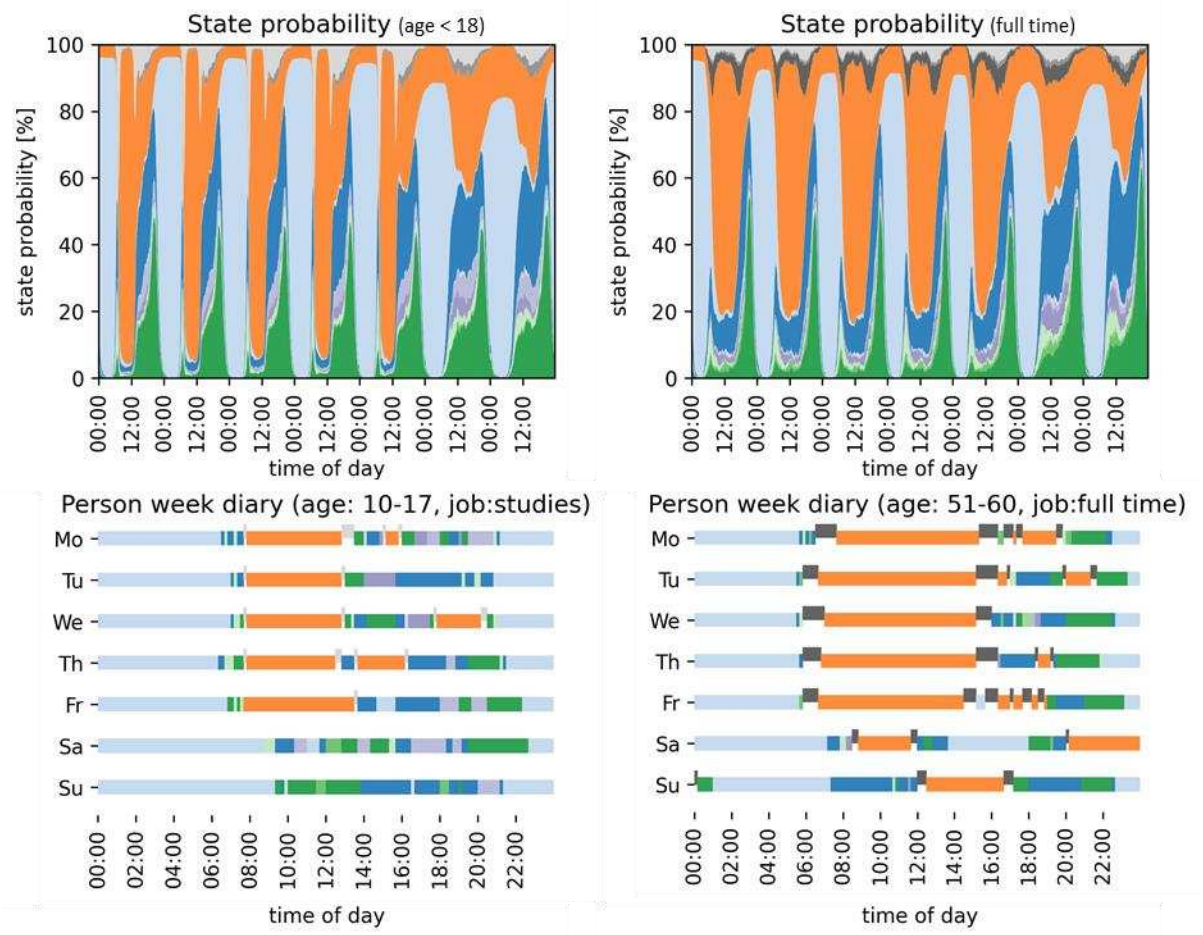


Figure 15: The top two figures represent the course of the aggregated state probability for 1,500 generated activity plans for persons under 18 years of age and for full time employees. The lower two representations are two exemplary activity plans for a person under the age of 18 and a full-time employee (A legend can be found in Figure 4).

5. Discussion

The results of Section 4.1 show that the Markov model used as a reference model is not able to record long-term dependencies in activity patterns and, due to the structure of the approach, is not able to adequately record the diversity in occupancy behavior. Consequently, synthetic activity schedules generated with Markov chains cannot be used to analyse occupancy behavior on an individual level and are only suitable for studies on an aggregated level. The approach presented in this paper combines weekly mobility data with a large sample size with high-resolution activity data with the help of new machine learning algorithms. The approach creates a new data basis which can be used for further analyses of home occupancy and mobility behavior. The profiles generated have similar stochastic properties as the empirically collected data on both the individual and the aggregated level.

By adequately capturing long-term dependencies in people's activities, the behavior of individual people can be reproduced. As a result, the data generated represent the basis for a variety of potential applications, one of which is the examination of potential charging periods of people with electric vehicles, assuming that electric mobility does not change mobility behavior. By combining the detailed mobility data with high-resolution activity data, a unique data basis is created which offers the possibility of consistently simulating the energy demand from personal mobility, the electrical demand for household devices and the heat demand for space heating and domestic hot water. Therefore, simultaneity effects in energy demand can be analysed based on one fundamental data set.

When analyzing such future developments, it should be taken into account that the data sets on which this work is based describe historical behavior (MOP: 2001-2017, TUD: 2001/02). Not taking into account the dynamics in people's behavioral habits could lead to significant errors, depending on the application. The energy sector includes many examples of innovations that have changed people's behavior for example, the internal combustion engine for transport and the development of ICT in recent decades. Hence ground-breaking/disruptive technologies could change the nature of the energy service demand itself (e.g. autonomous electric vehicles and smart home applications). In order to take into account temporal changes in behavior in the data set, the survey year of the respective sample could be provided as additional information in future studies. Furthermore, the data sets used differ in their temporal resolution, while the mobility data (MOP) are available in minute resolution, activities in the TUD are recorded in ten-minute resolution. The aggregation of the mobility data to a temporal resolution of 10 minutes can lead to distortions in short mobility states.

Through the use of machine learning approaches the assumption bias in the presented approach is low in comparison to e.g. utility-based stepwise regression approaches (Hilgert et al. 2017), therefore the developed approach is highly transferable. TUD data are collected uniformly in several European countries, but there are some differences in the design of the surveys. Some countries only provide activity time series for one weekday and one weekend day, which makes it harder to capture interday dependencies in activities. Longitudinal surveys of mobility behavior are not carried out in a harmonized way at the European level. However, similar mobility studies are available, for example in the UK and the Netherlands, which examine the mobility behavior over a whole week of a sample that is representative of the nation (Department for Transport 2020; Hoogendoorn-Lanser et al. 2015). The approach presented could therefore easily be applied to behavioral data in the UK and the Netherlands. Instead of training individual models for different countries, it would make more sense to implement the country information as a socio-demographic parameter in a transnational model in order to learn country-specific behavior and at the same time provide the model with a larger database for learning general behavioral relationships.

In this work, the focus was placed on the mapping of the mobility and activity behavior of individual persons and therefore no interpersonal relationships in the behavior of several individuals in a household were taken into account. However, the presented approach can and will be extended to represent household behavior in order to capture interpersonal relationships. Furthermore, only socio-demographic behavioral differences based on age and employment are currently taken into account in the model. Since the underlying data sets contain significantly more socio-demographic differentiations, an extension to include further socio-demographic characteristics is possible.

Since the training process is stopped before the presented models overfit, it can be stated that the models have learned the general stochastic relationships in human behavior and not simply learned the raw data sets by heart. This statement is supported by Figure 19 in the appendix, which describes the distribution of the minimum distances of a sample of data set a with all samples of data set b. The distribution of the minimum distances between the synthetic mobility schedules and the raw data is similar to the distribution of the minimum distances within the empirically collected data. However, even if the raw data used in this paper are already provided in anonymized form, it must be ensured that no information about individual samples in the empirical data is revealed by the synthetic data sets. Consequently, in follow-up work, prior to making the models presented in this paper available to the general public, algorithms from the field of "differential privacy" must be used to ensure that no information about individual samples is provided (Dwork and Roth 2014). Algorithms that ensure the privacy of individuals have been developed in recent years for deep learning applications (Abadi et al. 2016). Ensuring differential privacy is always accompanied by a loss of quality in the model, whereby this trade-off between quality and privacy can be clearly quantified by the so-called privacy budget.

6. Conclusion and Outlook

Over the past few years, many models have been published that aim to capture relationships in activity patterns to explain residential energy demand. Most of these models are different Markov variants or regression models that have a strong assumption bias and are therefore unable to capture complex long-term dependencies and the diversity in occupancy behavior. In this work it was shown that machine learning models from the field of natural language processing are able to capture long-term dependencies in mobility and activity patterns and at the same time adequately depict the diversity in behavior across the entire population. In a first step, two autoregressive models are presented which are able to recognize and reproduce weekly mobility patterns. In a second step, two imputation models are trained with time use data, which, based on the mobility information of individual people, enrich them with energy-related activities. Finally, the two models are combined to generate weekly activity plans. By combining an autoregressive generative model with an imputation model, the advantages of two data sets are combined and new data are generated which are beneficial for multiple use cases. One of which is the examination of flexibility potentials of individual households which is urgently needed for the integration of volatile renewable energy sources. Furthermore, metrics were introduced that enable activity profiles to be investigated in terms of intrapersonal and interpersonal variability. Based on these metrics, it is shown that the synthetically generated activity plans represent weekly mobility patterns and day-to-day dependencies of the energy-relevant activities with a high quality on an individual and aggregated level. The evaluation metrics show that LSTM and attention-based neural networks outperform existing approaches on an individual level by a large margin and at the same time have only slight deviations in the aggregated behavior.

Due to the availability of rich socio-demographic information in the two basic data sets, activity plans can be generated for different socio-demographic groups and can be used in future work to simulate consistent energy demand profiles from electric mobility, household devices and space heating. The approach developed provides the basis for making high-quality weekly activity data available to the general public without having to carry out complex application procedures. It was shown that the presented approach does not learn the training data by heart, however, it must be ensured that no private information about individuals is revealed by the model before the synthetic data can be provided to the community, which cannot be ensured at the current time. Therefore, in further work the model will be trained in a differential private way. Furthermore, the presented methodology can be trained with behavioral data from different European countries in order to develop a transnational model. Instead of individual behavior, household behavior could be learned to take interpersonal dependencies into account.

Acknowledgement

This work was supported by the Helmholtz Association under the Joint Initiative “Energy Systems Integration” (funding reference: ZT-0002) and was done during a research stay funded by the Centre for Research into Energy Demand Solutions (CREDS) at the University of Reading (UK). This work was supported by UKRI [grant numbers EP/R000735/1, EP/R035288/1 and EP/P000630/1].

7. References

Abadi, Martin; Chu, Andy; Goodfellow, Ian; McMahan, H. Brendan; Mironov, Ilya; Talwar, Kunal; Zhang, Li (2016): Deep Learning with Differential Privacy. In *CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318. DOI: 10.1145/2976749.2978318.

Aerts, D.; Minnen, J.; Glorieux, I.; Wouters, I.; Descamps, F. (2014): A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. In *Building and Environment* 75, pp. 67–78. DOI: 10.1016/j.buildenv.2014.01.021.

Bahdanau, Dzmitry; Cho, Kyunghyun; Bengio, Yoshua (2015): Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of International Conference on Learning Representations*. Available online at http://arxiv.org/pdf/1409.0473v7.

Bengio, Yoshua; Ducharme, Réjean; Vincent, Pascal; Jauvin, Christian (2003): A neural probabilistic language model. In *Journal of Machine Learning Research (3)*, pp. 1137–1155.

Bottaccioli, Lorenzo; Di Cataldo, Santa; Acquaviva, Andrea; Patti, Edoardo (2019): Realistic Multi-Scale Modeling of Household Electricity Behaviors. In *IEEE Access* 7, pp. 2467–2489. DOI: 10.1109/ACCESS.2018.2886201.

Bowman, John (1998): The Day Activity Schedule Approach to Travel Demand Analysis. Dissertation. Cambridge, Massachusetts.

Brown, Tom B.; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared; Dhariwal, Prafulla et al. (2020): Language Models are Few-Shot Learners, 5/28/2020. Available online at http://arxiv.org/pdf/2005.14165v4.

Caccia, Massimo; Caccia, Lucas; Fedus, William; Larochelle, Hugo; Pineau, Joelle; Charlin, Laurent (2020): Language GANs Falling Short. In *Proceedings of the Seventh International Conference on Learning Representation*. Available online at https://openreview.net/pdf?id=r1lOgyrKDS.

Chen, Zhenghua; Jiang, Chaoyang; Xie, Lihua (2018): Building occupancy estimation and detection. A review. In *Energy and Buildings* 169, pp. 260–270. DOI: 10.1016/j.enbuild.2018.03.084.

Department for Transport (2020): 2019 National Travel Survey.

Destatis (2006): Zeitbudgeterhebung: Aktivitäten in Stunden und Minuten nach Geschlecht, Alter und Haushaltstyp. Zeitbudgets - Tabellenband I. 2001/2002. Wiesbaden. Available online at https://www.statistischebibliothek.de/mir/receive/DEMonografie_mods_00003054.

Devlin, Jacob; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina (2018): BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Available online at http://arxiv.org/pdf/1810.04805v2.

Dwork, Cynthia; Roth, Aaron (2014): The Algorithmic Foundations of Differential Privacy. In *FNT in Theoretical Computer Science* 9 (3-4), pp. 211–407. DOI: 10.1561/0400000042.

Elman, Jeffrey L. (1990): Finding Structure in Time. In *Cognitive Science* 1990 (14), pp. 179–211.

Eurostat (2000): Harmonized European Time of Use Survey.

Fischer, David; Härtl, Andreas; Wille-Haussmann, Bernhard (2015): Model for electric load profiles with high time resolution for German households. In *Energy and Buildings* 92, pp. 170–179. DOI: 10.1016/j.enbuild.2015.01.058.

Flett, Graeme; Kelly, Nick (2016): An occupant-differentiated, higher-order Markov Chain method for prediction of domestic occupancy. In *Energy and Buildings* 125, pp. 219–230. DOI: 10.1016/j.enbuild.2016.05.015.

Goodfellow, Ian; Jean Pouget-Abadie; Mehdi Mirza; Bing Xu; David Warde-Farley; Sherjil Ozair et al. (2014): Generative Adversarial Nets. In *Advances in neural information processing systems, pp 2672–2680*.

Hilgert, Tim; Heilig, Michael; Kagerbauer, Martin; Vortisch, Peter (2017): Modeling Week Activity Schedules for Travel Demand Models. In *Transportation Research Record* 2666 (1), pp. 69–77. DOI: 10.3141/2666-08.

Hochreiter, S.; Schmidhuber, J. (1997): Long Short-Term Memory. Available online at https://www.mitpressjournals.org/doi/10.1162/neco.1997.9.8.1735, checked on 8/13/2020.136Z.

Hoogendoorn-Lanser, Sascha; Schaap, Nina T.W.; OldeKalter, Marie-José (2015): The Netherlands Mobility Panel: An Innovative Design Approach for Web-based Longitudinal Travel Data Collection. In *Transportation Research Procedia* 11, pp. 311–329. DOI: 10.1016/j.trpro.2015.12.027.

Kalchbrenner, Nal; Grefenstette, Edward; Blunsom, Phil (2014): A Convolutional Neural Network for Modelling Sentences. Available online at http://arxiv.org/pdf/1404.2188v1.

Kaschub, Thomas (2017): Batteriespeicher in Haushalten unter Berücksichtigung von Photovoltaik, Elektrofahrzeugen und Nachfragesteuerung. Dissertation.

McKenna, Eoghan; Krawczynski, Michal; Thomson, Murray (2015): Four-state domestic building occupancy model for energy demand simulations. In *Energy and Buildings* 96, pp. 30–39. DOI: 10.1016/j.enbuild.2015.03.013.

Mikolov, Tomas; Chen, Kai; Corrado, Greg; Dean, Jeffrey (2013): Efficient Estimation of Word Representations in Vector Space. Available online at http://arxiv.org/pdf/1301.3781v3.

Paardekooper, Susana; Lund, Rasmus Søgaard; Mathiesen, Brian Vad; Chang, Miguel (2018): Heat Roadmap Europe 4. Quantifying the Impact of Low-Carbon Heating and Cooling Roadmaps.

Pflugradt, Noah Daniel (2016): Modellierung von Wasser und Energieverbräuchen in Haushalten. Available online at http://nbn-resolving.de/urn:nbn:de:bsz:ch1-qucosa-209036, checked on 4/12/2018.

Ramírez-Mendiola, José Luis; Grünewald, Philipp; Eyre, Nick (2019): Residential activity pattern modelling through stochastic chains of variable memory length. In *Applied Energy* 237, pp. 417–430. DOI: 10.1016/j.apenergy.2019.01.019.

Richardson, Ian; Thomson, Murray; Infield, David (2008): A high-resolution domestic building occupancy model for energy demand simulations. In *Energy and Buildings* 40 (8), pp. 1560–1566. DOI: 10.1016/j.enbuild.2008.02.006.

Richardson, Ian; Thomson, Murray; Infield, David; Clifford, Conor (2010): Domestic electricity use. A high-resolution energy demand model. In *Energy and Buildings* 42 (10), pp. 1878–1887. DOI: 10.1016/j.enbuild.2010.05.023.

Shove, Elizabeth; Pantzar, Mika; Watson, Matt (2012): The Dynamics of Social Practice: Everyday Life and How it Changes: SAGE Publications Ltd.

Steemers, Koen; Yun, Geun Young (2009): Household energy consumption. A study of the role of occupants. In *Building Research & Information* 37 (5-6), pp. 625–637. DOI: 10.1080/09613210903186661.

Sutskever, Ilya (2013): Training Recurrent Neural Networks. Dissertation.

Sutskever, Ilya; Vinyals, Oriol; Le V, Quoc (2014): Sequence to Sequence Learning with Neural Networks. Available online at http://arxiv.org/pdf/1409.3215v3.

Tjaden, Tjarko; Bergner, Joseph; Weniger, Johannes; Quaschning, Volker (2015): Repräsentative elektrische Lastprofile für Wohngebäude in Deutschland auf 1-sekündiger Datenbasis.

Torriti, Jacopo (2014): A review of time use models of residential electricity demand. In *Renewable and Sustainable Energy Reviews* 37, pp. 265–272. DOI: 10.1016/j.rser.2014.05.034.

826 Torriti, Jacopo (2017): Understanding the timing of energy demand through time use data. Time of the
827 day dependence of social practices. In *Energy Research & Social Science* 25, pp. 37–47. DOI:
828 10.1016/j.erss.2016.12.004.

829 Transport & Environment (2020): Recharge EU. How many charge points will Europe and its member
830 states need in the 2020s. With assistance of William Todts. Available online at
831 https://www.transportenvironment.org/sites/te/files/publications/01%202020%20Draft%20TE%20In
832 frastructure%20Report%20Final.pdf.

833 Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N. et al.
834 (2017): Attention Is All You Need. Available online at http://arxiv.org/pdf/1706.03762v5.

835 Walker, Gordon (2014): The dynamics of energy demand: Change, rhythm and synchronicity. In *Energy*
836 *Research & Social Science* 1, pp. 49–55. DOI: 10.1016/j.erss.2014.03.012.

837 Weiß, Christine; Chlond, Bastian; Hilgert, Tim; Vortisch, Peter (2016): Deutsches Mobilitätspanel (MOP)
838 - wissenschaftliche Begleitung und Auswertungen, Bericht 2014/2015. Alltagsmobilität und
839 Fahrleistung, checked on 10/22/2019.

840 Wilke, Urs (2013): Probabilistic Bottom-up Modelling of Occupancy and Activities to Predict Electricity
841 Demand in Residential Buildings. PHD Thesis.

842 Williams, R. J.; Zipser, D. (1989): A learning algorithm for continually running fully recurrent neural
843 networks. In *Neural computation* (1(2)), 270–280.

844 Wolf, Thomas; Debut, Lysandre; Sanh, Victor; Chaumond, Julien; Delangue, Clement; Moi, Anthony et
845 al. (2020): HuggingFace's Transformers: State-of-the-art Natural Language Processing. Available online
846 at http://arxiv.org/pdf/1910.03771v5.

847 Wu, Yonghui; Schuster, Mike; Chen, Zhifeng; Le V, Quoc; Norouzi, Mohammad; Macherey, Wolfgang
848 et al. (2016): Google's Neural Machine Translation System: Bridging the Gap between Human and
849 Machine Translation. Available online at http://arxiv.org/pdf/1609.08144v2.

850 Yamaguchi, Y.; Yilmaz, S.; Prakash, N.; Firth, S. K.; Shimoda, Y. (2018): A cross analysis of existing
851 methods for modelling household appliance use. In *Journal of Building Performance Simulation* (5673),
852 pp. 1–20. DOI: 10.1080/19401493.2018.1497087.

853 Yilmaz, Selin; Firth, Steven K.; Allinson, David (2017): Occupant behaviour modelling in domestic
854 buildings: the case of household electrical appliances. In *Journal of Building Performance Simulation*
855 10 (5-6), pp. 582–600. DOI: 10.1080/19401493.2017.1287775.

856 Young, Tom; Hazarika, Devamanyu; Poria, Soujanya; Cambria, Erik (2017): Recent Trends in Deep
857 Learning Based Natural Language Processing. Available online at http://arxiv.org/pdf/1708.02709v8.

858 Yu, Lantao; Zhang, Weinan; Wang, Jun; Yu, Yong (2017): SeqGAN: Sequence Generative Adversarial
859 Nets with Policy Gradient. Available online at https://www.nature.com/articles/nature16961.pdf.

860 Zumkeller, D.; Chlond, B. (2009): Dynamics of Change: Fifteen-Year German Mobility Panel. Presented
861 at 88th Annual Meeting of the Transportation Research Board. Washington, D.C., 2009.

862

8. Appendix

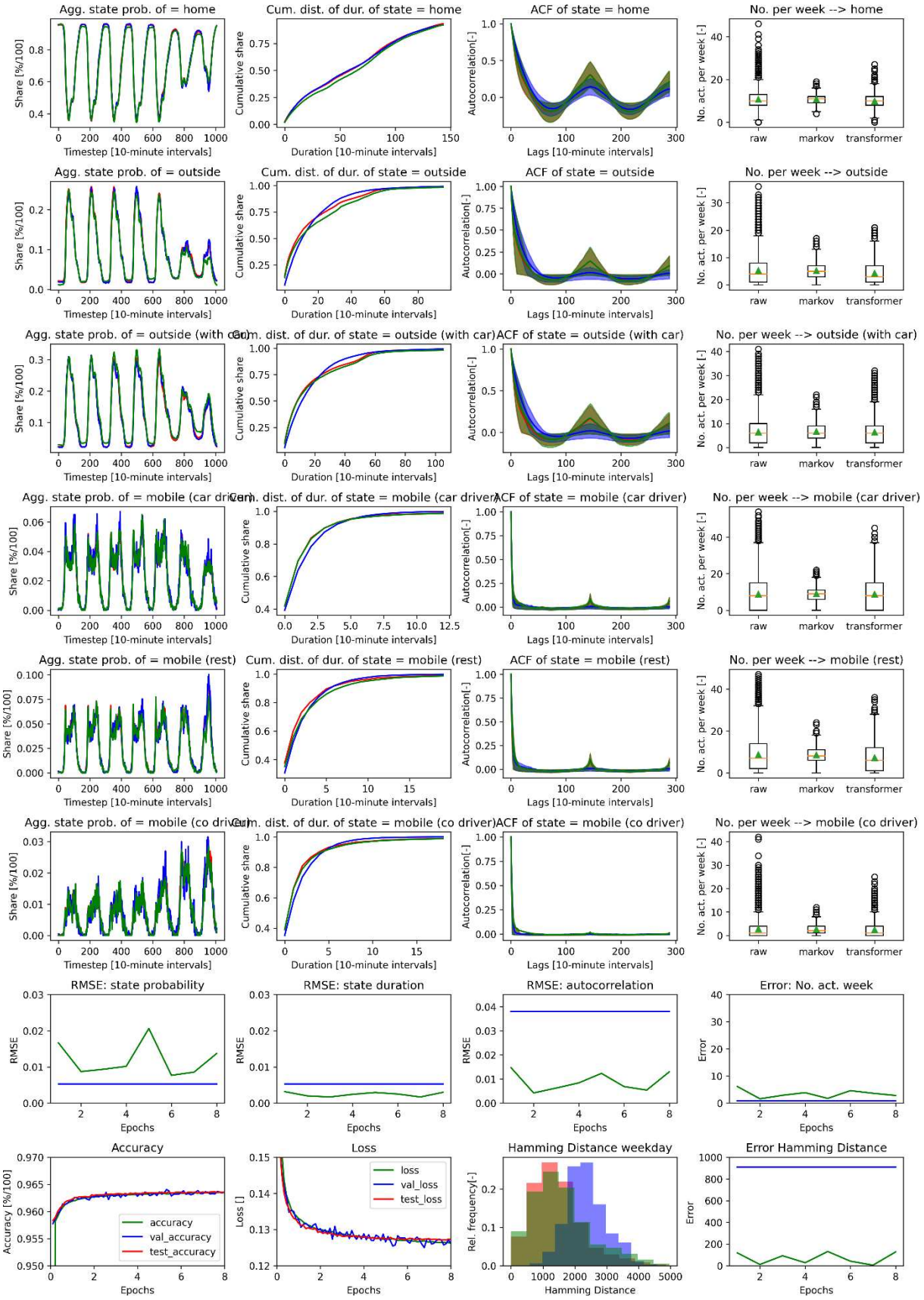


Figure 16: Comparison of all metrics and all states for the mop data (red), the attention based autoregressive model described in Table 3 (no. 3) (green) and a first order Markov model (blue). The mobility schedule specific metrics of the attention based model are calculated based on the model weights after epoch 7.

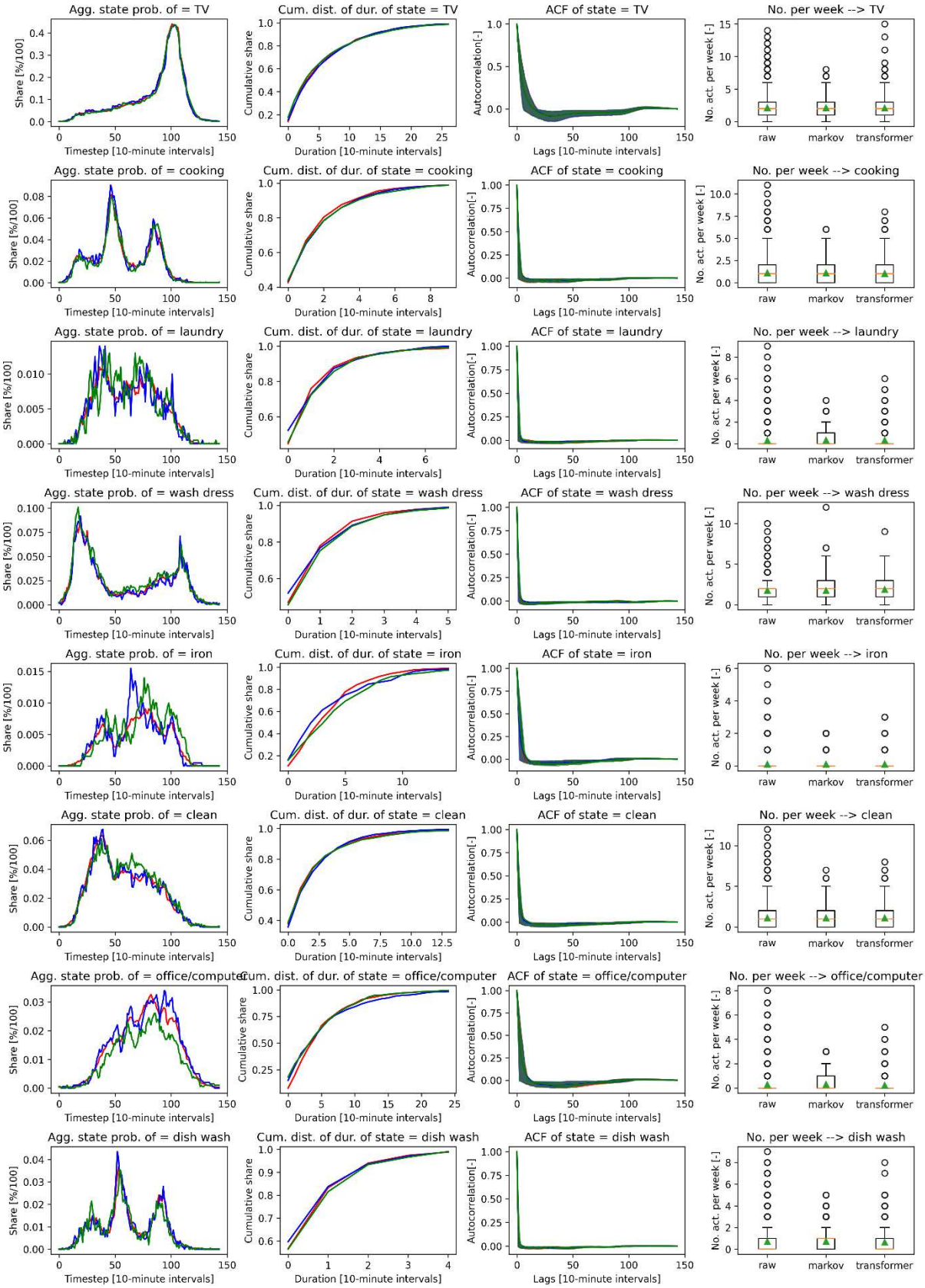


Figure 17: Part a: Comparison of all metrics and all states for the TUD data (red), the attention based imputation model described in Table 5 (model no. 6) (green) and a first order Markov model (blue – no imputation model). The mobility schedule specific metrics of the attention based model are calculated based on the model weights after epoch 37. The autocorrelation graphs were calculated based on single days.

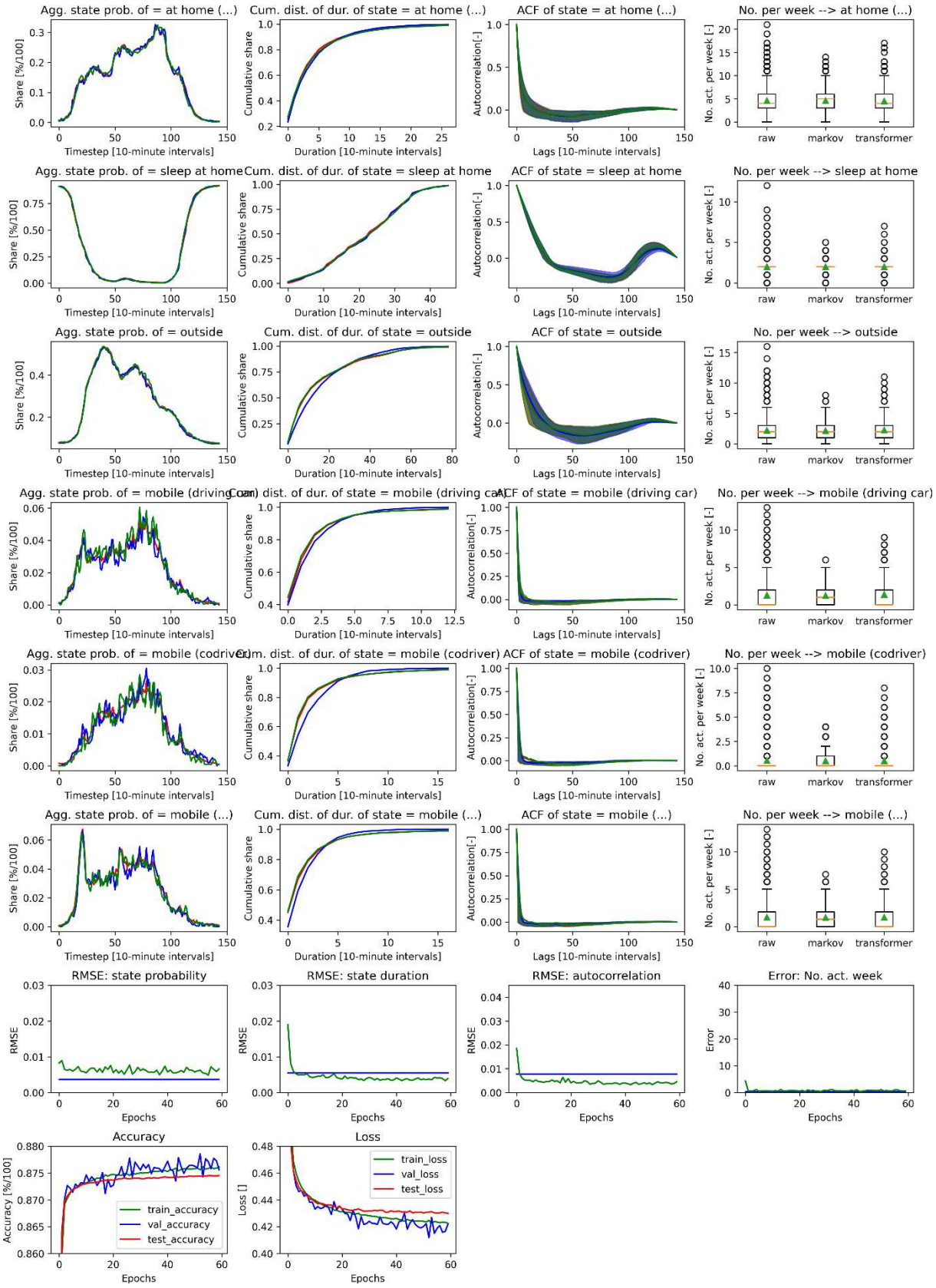


Figure 18: Part b: Comparison of all metrics and all states for the TUD data (red), the attention based imputation model described in Table 5 (model no. 6) (green) and a first order Markov model (blue – no imputation model). The mobility schedule specific metrics of the attention based model are calculated based on the model weights after epoch 37. The autocorrelation graphs were calculated based on single days. Furthermore, the course of the model loss and accuracy is visualized.

Table 2: Hyperparameter configurations and model metrics for the LSTM based autoregressive model. Metrics were calculated based on a sample size of N=2,000. Furthermore, a mean standard error due to the sample size of 2,000 is given.

No.	LSTM units/ Learning rate/ Batch size/ Dense neurons	Sp rmse [%]	Sd rmse [%]	Ac rmse [%]	Na mae []	Hd mae []	Cross-entropy Loss	Accuracy [%]	CV Epochs
1	512/0.0005/512/32	0.99	0.13	0.71	1.11	144	0.133	96.27	14
2	128/0.0005/512/32	1.03	0.17	1.65	1.57	423	0.142	96.12	8
3	512/0.001/512/32	1.05	0.18	0.66	3.04	235	0.131	96.30	11
4	512/0.0005/64/32	1.27	0.22	0.89	3.39	114	0.134	96.26	3
5	512/0.0005/512/64	0.90	0.18	0.80	0.67	98	0.131	96.29	17
6	512/0.0005/256/32	0.90	0.13	0.60	1.85	83	0.131	96.29	11
7	2x256/0.001/512/32	0.69	0.14	0.95	2.08	120	0.131	96.29	12
8	2x256/0.0005/256/32	0.97	0.19	0.63	3.61	1.5	0.132	96.28	10
Standard error (N=2,000)		0.52	0.09	0.24	0.6	13	-	-	-

Table 3: Hyperparameter configurations and model metrics for the attention based autoregressive model. 2xh means that two attention heads are used (see (Vaswani et al. 2017)).

No.	Transformer layers/ D_model/ Learning rate/ Batch size	Sp rmse	Sd rmse	Ac rmse	Na mae	Hd mae	Cross-entropy Loss	Accuracy	CV Epochs
1	1/64/0.001/64	0.83	0.31	1.32	2.96	244	0.14	95.95	9
2	4/64/0.001/64	0.91	0.16	0.70	2.53	33	0.128	96.34	15
3	**8/64/0.001/64**	**0.86**	**0.17**	**0.54**	**3.6**	**5**	**0.127**	**96.36**	**7**
4	4/64/0.001/128	0.95	0.22	0.54	3.28	44	0.130	96.29	3
5	4/128/0.001/128	0.89	0.24	0.59	3.60	9	0.128	96.33	6
6	4/64/0.0005/64	0.86	0.18	0.48	4.78	6	0.128	96.33	20
7	2(2xh)/64/0.001/64	0.97	0.22	0.60	6.33	74	0.129	96.31	11
8	4(2xh)/64/0.001/64	1.20	0.20	0.42	4.52	126	0.127	96.35	8
Standard errors (N=2,000)		0.52	0.09	0.24	0.6	13	-	-	-

Table 4: Hyperparameter configurations and model metrics for the BiLSTM based imputation model. Metrics were calculated based on a sample size of N=2,000 diary days. Furthermore, a mean standard error due to the sample size of 2,000 diary days is given.

No.	LSTM units/ D_model/ Learning rate/ Batch size	Sp rmse	Sd rmse	Ac rmse	Na mae	Cross-entropy Loss	Accuracy	CV Epochs
1	64/32/0.001/64	0.70	0.27	0.36	0.88	0.434	87.48	21
2	128/32/0.001/64	0.74	0.28	0.44	0.86	0.435	87.36	11
3	256/32/0.001/64	0.60	0.26	0.37	0.59	0.432	87.46	9
4	128/64/0.001/128	0.75	0.26	0.42	0.96	0.432	87.54	13
5	128/32/0.001/128	0.71	0.42	0.48	0.98	0.433	87.44	11
6	128/32/0.0005/128	0.64	0.28	0.43	1.27	0.434	87.48	12
7	64/32/0.0005/128	0.60	0.30	0.38	0.62	0.434	87.39	33
8	64/32/0.0005/64	0.62	0.34	0.44	0.82	0.434	87.43	33
Standard errors (N=2,000)		0.40	0.19	0.24	0.33	-	-	-

889

890 *Table 5: Hyperparameter configurations and model metrics for the attention based imputation model. Metrics were calculated*
891 *based on a sample size of N=2,000 diary days. Furthermore, a mean standard error due to the sample size of 2,000 diary days*
892 *is given.*

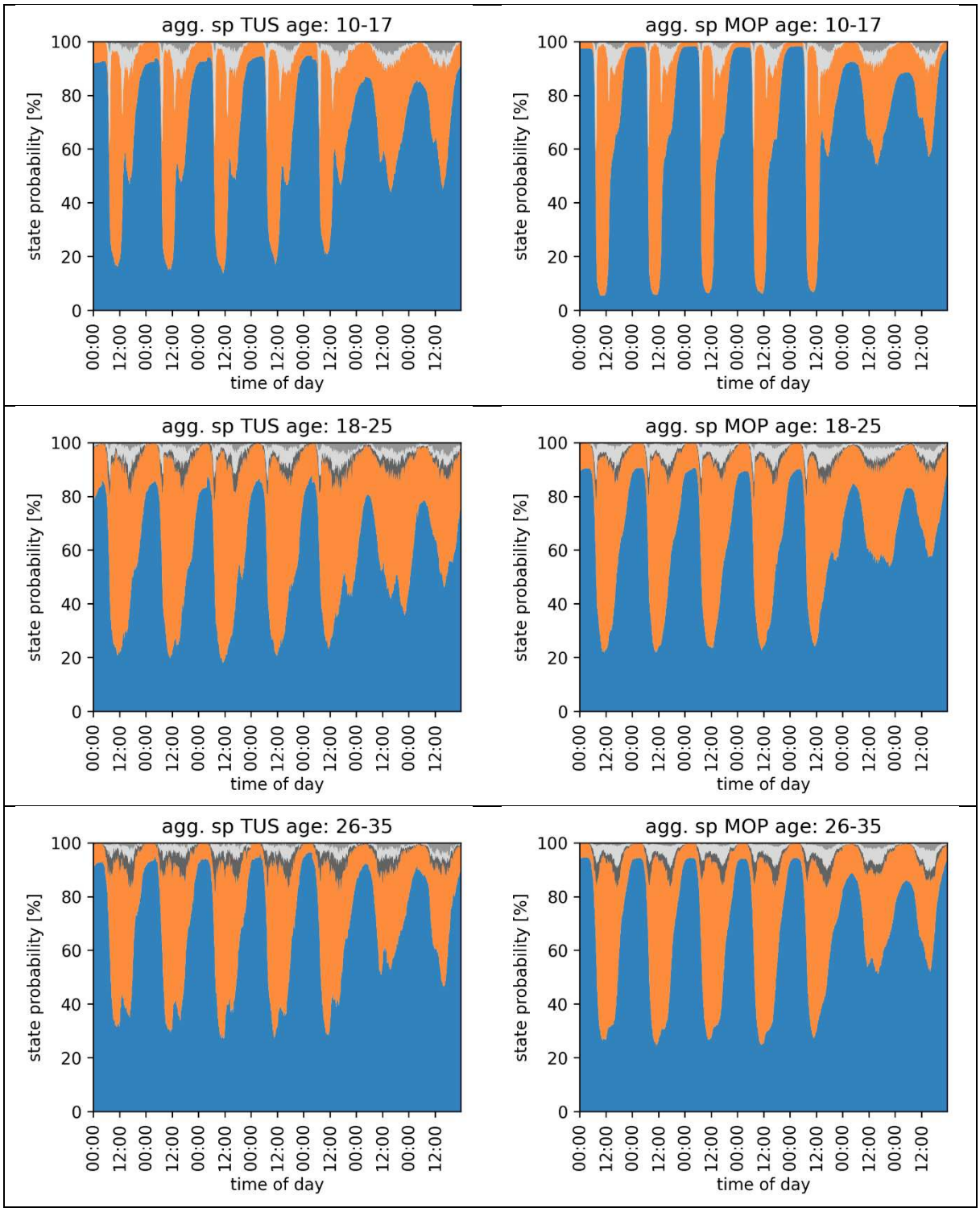
No.	Transformer layers/ D_model/ Learning rate/ Batch size	Sp rmse	Sd rmse	Ac rmse	Na mae	Cross-entropy Loss	Accuracy	CV Epochs
1	1/64/0.001/256	0.58	0.39	0.50	0.50	0.469	86.97	158
2	4/64/0.001/256	0.58	0.39	0.44	0.62	0.436	87.32	22
3	4/64/0.001/64	0.57	0.38	0.36	0.90	0.436	87.35	8
4	4/64/0.001/128	0.63	0.39	0.46	1.05	0.438	87.31	12
5	4/64/0.0005/64	0.59	0.36	0.39	0.72	0.435	87.35	10
6	**4/64/0.0005/128**	**0.49**	**0.39**	**0.39**	**0.66**	**0.431**	**87.41**	**37**
7	4/14/0.0005/64	1.27	0.54	0.72	1.42	0.458	87.14	46
8	4/14/0.0005/128	0.84	0.60	0.61	0.65	0.459	87.14	47
Standard errors (N=2,000)		0.40	0.19	0.24	0.33	-	-	-

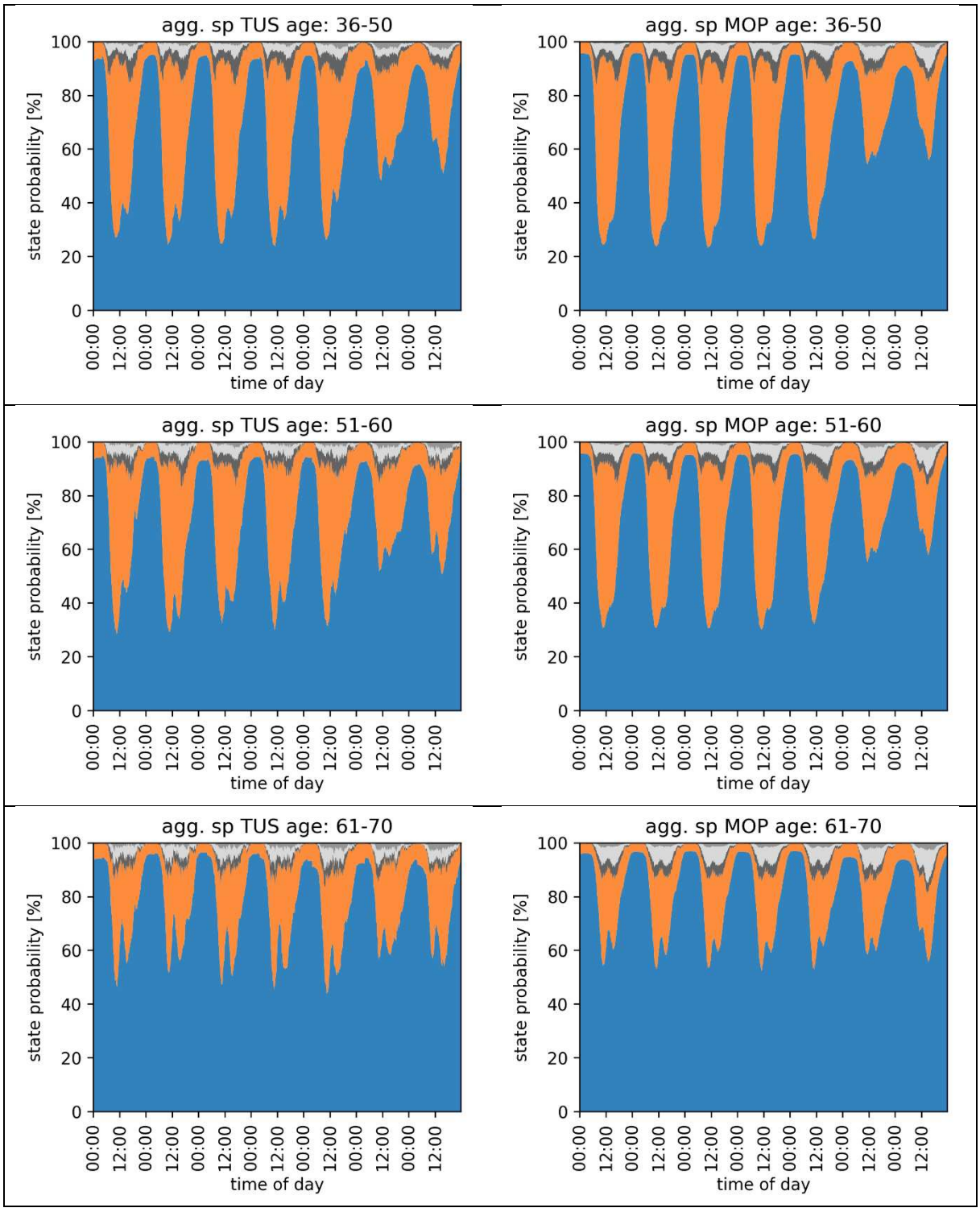
893

894 *Table 6: Comparative presentation of the socio-demographic composition of the MOP and TUD data sets. The calculated rmse*
895 *of the aggregated state probabilities are calculated on the basis of the five aggregated states (home, outside, mobile (car*
896 *driver), mobile (co driver), mobile (rest)). For the calculation of the rmse between the synthetic profiles and the MOP and TUD*
897 *data, synthetic data with the same socio-demographic characteristics as in the comparison data sets were generated.*

Age	**<18**	**<26**	**<36**	**<51**	**<61**	**<71**	**>=71**
Samples MOP	1971 (7.4%)	1430 (5.4%)	2288 (8.6%)	6107 (22.9%)	5132 (19.3%)	5809 (21.8%)	3873 (14.6%)
Samples TUD	2169 (18.2%)	1106 (9.3%)	1140 (9.6%)	4080 (34.2%)	1654 (13.9%)	1167 (9.8%)	494 (4.1%)
rmse sp MOP/TUD	4.0%	3.8%	2.3%	1.9%	2.2%	2.7%	2.8%
rmse sp syn./MOP	1.7%	1.6%	1.3%	0.9%	1.1%	0.7%	0.9%
rmse sp syn./TUD	3.9%	4.2%	2.1%	1.7%	1.9%	2.6%	2.7%
Job	**-**	**Full time**	**Part time**	**Students**	**Training**	**No job**	**Pensioner**
Samples MOP	212 (0.8%)	8853 (33.3%)	3627 (13.6%)	2759 (10.4%)	489 (1.8%)	2052 (7.7%)	8618 (32.4%)
Samples TUD	-	3938 (33.0%)	2599 (21.8%)	2214 (18.6%)	375 (3.1%)	1184 (9.9%)	1611 (13.5%)
rmse sp MOP/TUD	-	2.2%	2.5%	2.9%	3.8%	2.4%	2.4%
rmse sp syn./MOP	2.8%	1.1%	1.1%	1.3%	4.0%	1.0%	0.7%
rmse sp syn./TUD	-	2.1%	2.0%	3.6%	4.69%	2.1%	2.1%
rmse sp MOP/TUD (entire sample)				2.9%			
rmse sp syn./TUD (entire sample)				1.8%			
rmse sp syn./MOP (entire sample)				0.7%			

898 *Table 7: Comparative representation of the aggregated state probabilities of the TUD and MOP data sets for population*
899 *groups with different ages.*




agg. sp TUS age: 36-50
agg. sp MOP age: 36-50
agg. sp TUS age: 51-60
agg. sp MOP age: 51-60
agg. sp TUS age: 61-70
agg. sp MOP age: 61-70
state probability [%]
time of day
0
20
40
60
80
100
00:00
12:00

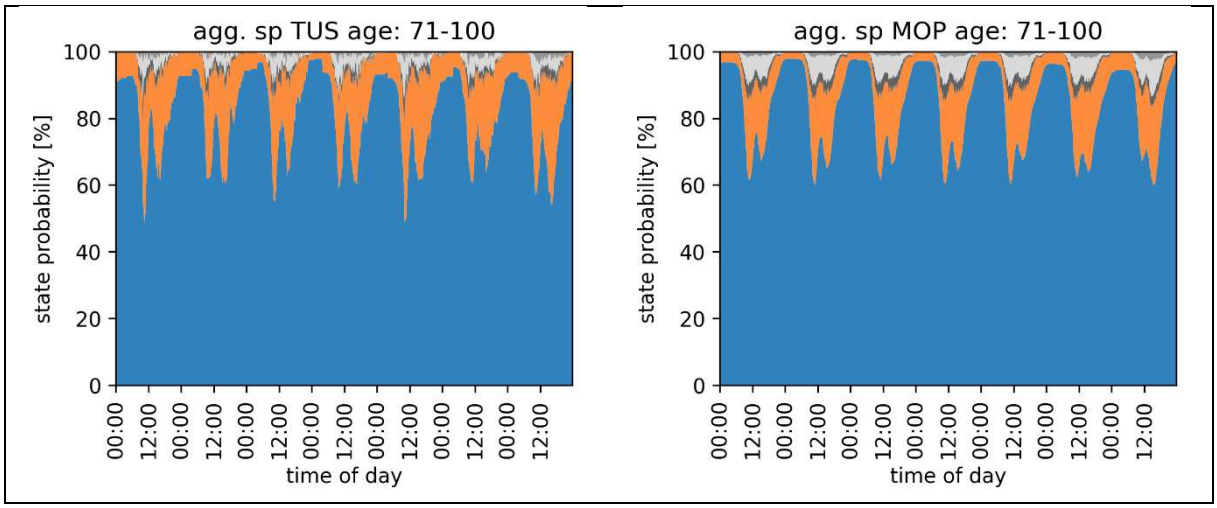
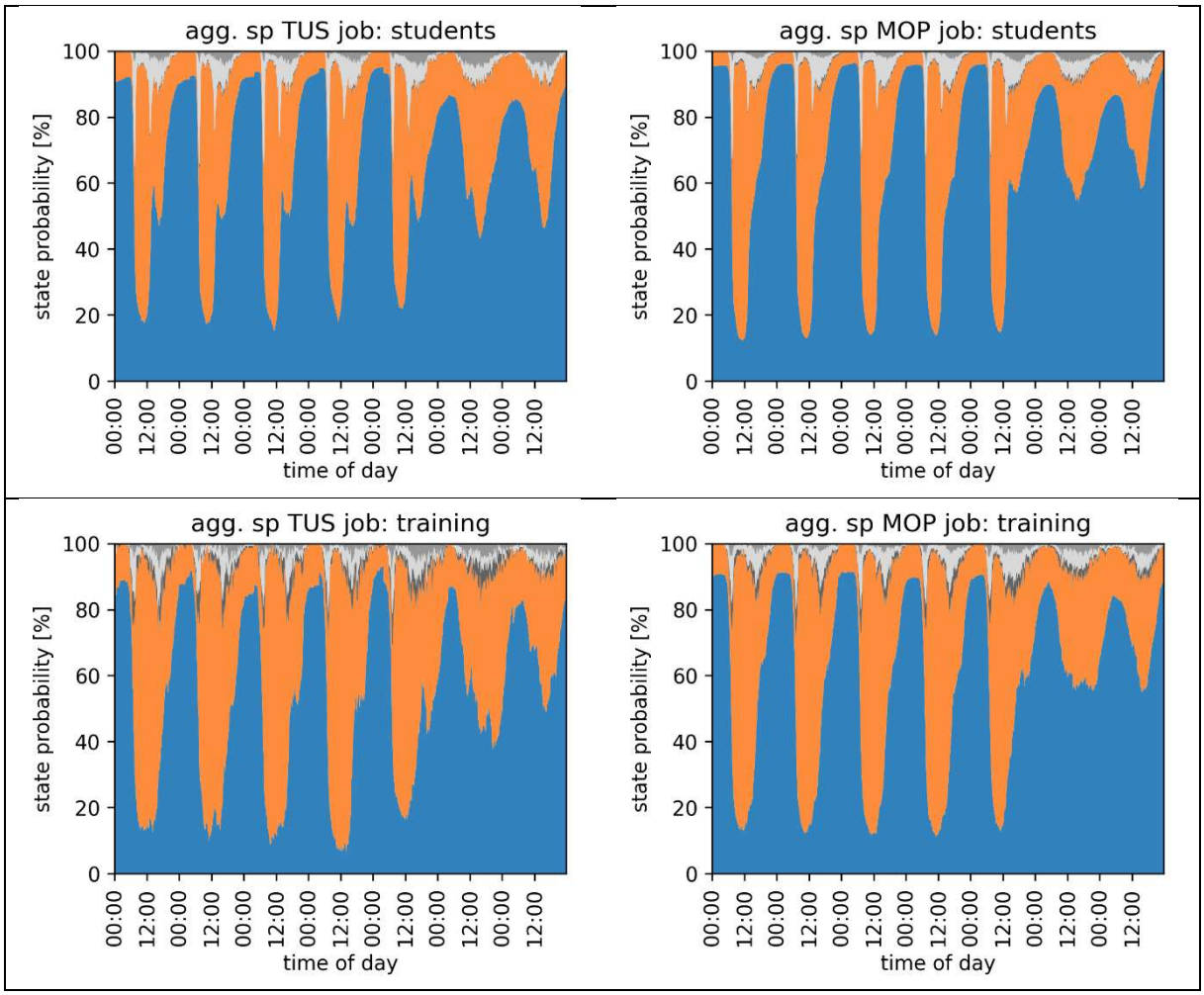
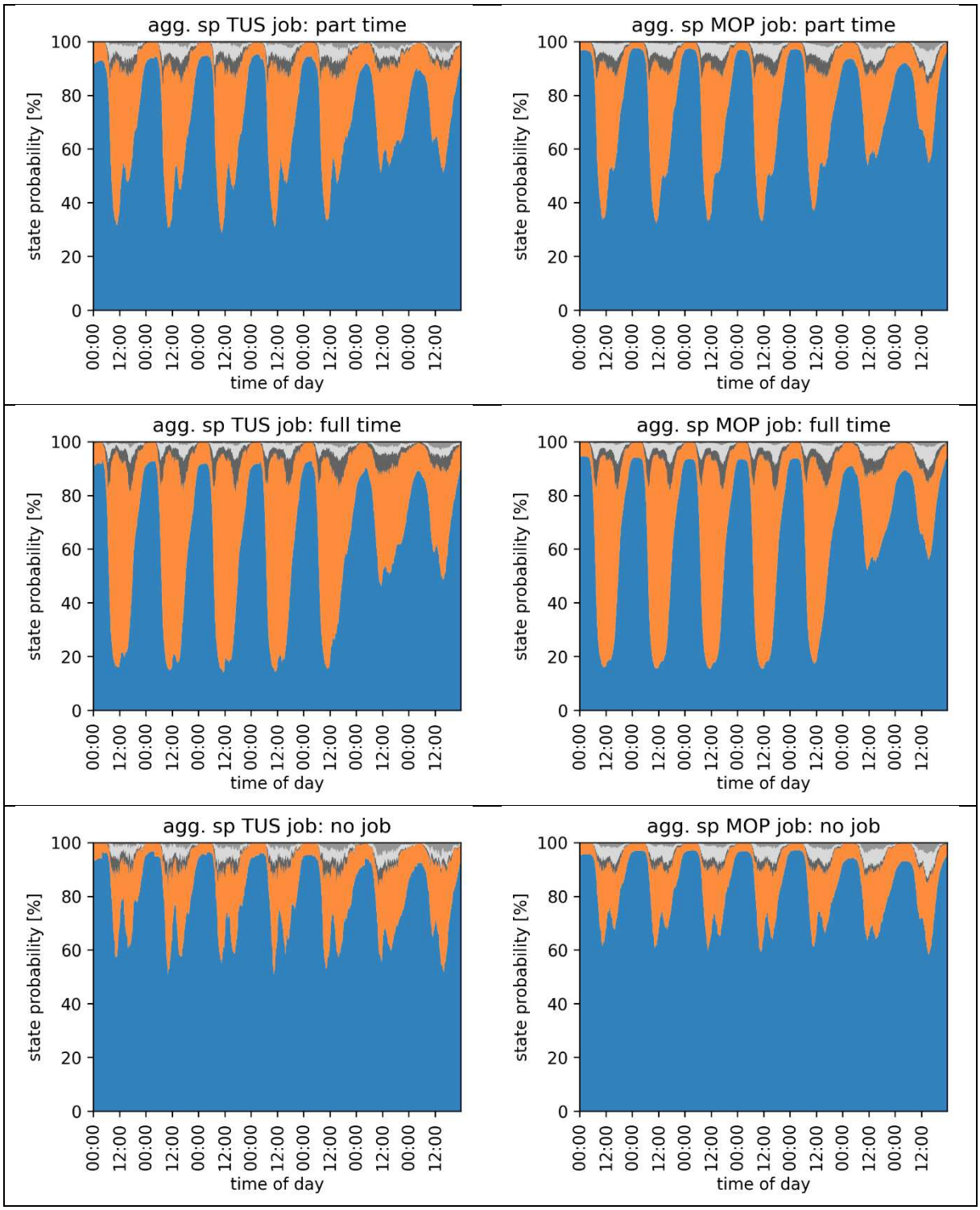
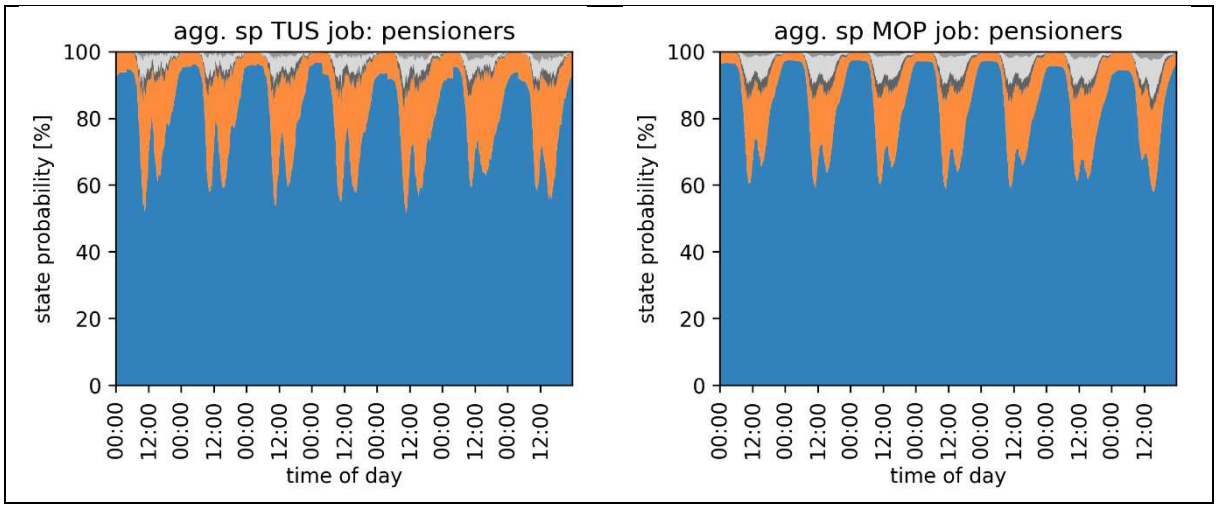


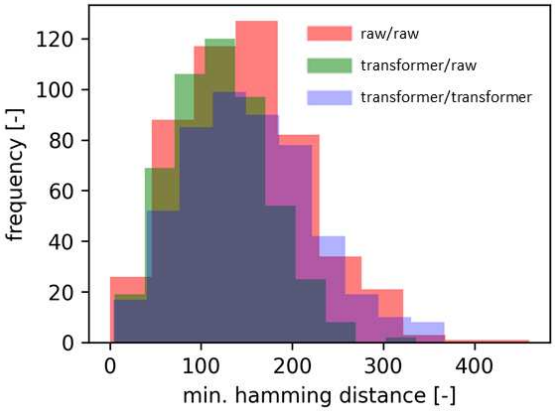
Table 8: Comparative representation of the aggregated state probabilities of the TUD and MOP data sets for population groups with different occupations.




agg. sp TUS job: part time
agg. sp MOP job: part time
agg. sp TUS job: full time
agg. sp MOP job: full time
agg. sp TUS job: no job
agg. sp MOP job: no job
state probability [%]
time of day
0
20
40
60
80
100
00:00
12:00



903



904

905 *Figure 19: Distribution of the minimum Hamming distances of the samples from dataset a (sample size N = 500) to the*
906 *samples in dataset b (dataset a/ dataset b)*

907

Declaration of interests

☒ 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.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: