Data-driven models for energy advising leading to behavioural changes in residences

Liisberg, Jon Anders Reichert

Publication date:
2019

Document Version
Publisher's PDF, also known as Version of record

Citation (APA):
Liisberg, J. A. R. (2019). *Data-driven models for energy advising leading to behavioural changes in residences*. Technical University of Denmark.

Data-driven models for energy advising leading to behavioural changes in residences

Jon A. R. Liisberg

Kongens Lyngby 2019

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

Summary (English)

Given the smart-meter roll-out in all of Europa and the implementation of the supplier-centric DataHub in Denmark since 2016, there is a great potential for increasing awareness among residential electricity consumers on their consumption. The smart-meter data is usually in hourly temporal resolution, available up-to one year in the past.

This thesis deals with feedback/advice to residential electricity consumers, based on data-driven models using the consumers own individual smart-meter data. Beside data-driven feedback a game was developed as a study on gamified interaction of electricity consumption.

Models for medium to long-term prediction of daily individual residential electricity consumption were developed. The models were applied in large-scale as a monitoring tool and presented to the consumers through the app Watts. The effect on electricity consumption of the feedback provided by the prediction model through the app Watts was evaluated, showing a decreasing effect for active users.

Methods for disaggregating hourly electricity consumption were also developed using Hidden Markov models (HMM). The states of the HMM are described in accordance to appliance-related activities and shows potential for probabilistic short-term load forecasting. These methods are yet to be applied in large-scale.

Resumé (Danish)

Med udrulningen af *smart-metre* i hele Europa og implementeringen af DataHubben for elforbrug i Danmark siden 2016, er der et stort potentiale for at øge bevidstheden hos el forbrugere omkring deres forbrug. *Smart-meter* dataene er normalt time målinger, der er tilgængelige op til et år tilbage i tiden.

Denne afhandling beskæftiger sig med feedback/rådgivning til el forbrugere, baseret på data-drevne modeller, der bruger forbrugerens egne individuelle *smart-meter* data. Udover data-dreven feedback blev et spil udviklet for at undersøge *gamified* interaktion med elforbrug.

Modeller til mellemlange og langsigtede prædiktioner af det daglige individuelle elforbrug blev udviklet. Modellerne blev anvendt i stor skala som et overvågningsværktøj og præsenteret til forbrugerne via appen Watts. Effekten på elforbruget på baggrund af feedbacket fra prædiktionsmodellen, via appen Watts, blev analyseret. Dette viste en faldende effekt for aktive brugere.

Metoder til disaggregering af time målinger på elforbrug blev også udviklet, ved hjælp af *Hidden Markov-modeller*. Modellernes tilstande blev beskrevet i overensstemmelse med apparatrelaterede aktiviteter i husstanden og viser potentiale for sandsynlighedsfordelt kort-tids fremskrivning af elforbruget. Disse metoder anvendes endnu ikke i stor skala.

Preface

This thesis was prepared at the Department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU) in partial fulfilment of the requirements for acquiring a Ph.D. degree.

This thesis addresses the development of models for medium to long-term prediction and disaggregation of individual residential electricity consumption. Methods are developed for large-scale application of the prediction models and an evaluation study considering the effect of the feedback provided by the prediction model. Furthermore, a study on gamified interaction of electricity consumption is addressed.

This thesis consists of a summary report, three research papers and three technical reports, detailing the research conducted over the period of June 2016 to June 2019.

Kgs. Lyngby, 14-June-2019

Jon A. R. Liisberg

Acknowledgements

I would like to express my gratitude to my supervisors at DTU Jan Kloppenborg Møller, Peder Bacher and Henrik Madsen for the useful discussions, and engagement through the process of my PhD project.

A special thank is going to my supervisor at SEAS-NVE, Anders Spur Hansen and the rest of my colleagues in DIDI, for making the project a reality and applying my research in Watts.

Thanks also goes to all of my colleagues at DynSys.

Particular thanks go to Joakim Widén and Joakim Munkhammar for the collaboration and for my visit to the Built Environment Energy Systems Group at Uppsala University.

Finally, I would like to give thanks to Daniel Cermak-Sassenrath from IT University of Copenhagen for the introduction to game development, the many discussions on gamification and guidance throughout the game development process.

This work was partly funded by Innovation Fund Denmark (IFD) under File No. 5189-00045B.

List of Publications

Scientific Research Publications in this Thesis

A J. Liisberg, J. Møller, P. Bacher. “Models for long-term baseline prediction of daily electricity consumption in individual households”. Submitted to *Applied Energy*.

B J. Liisberg, J. Møller, P. Bacher. “Extensions to the Recursive Least Squares modelling scheme. Self-tuning control and Box-Cox transformation”. DTU Technical Report 2019-xx.

C J. Liisberg, J. Møller, H. Bloem, J. Cipriano, G. Mor and H. Madsen. “Hidden Markov Models for indirect classification of occupant behaviour”. *Sustainable Cities and Society* 27 (2016): 83-98.

D J. Liisberg, J. Møller, P. Bacher, J. MunkHammar, J. Widén and H. Madsen. “Disaggregation using Continuous Time Hidden Markov Models applied on hourly electricity consumption”. Submitted to *Applied Energy*.

E J. Liisberg. “Gamification of electricity consumption, PowerMon”. DTU Technical Report 2019-xx.

F J. Liisberg, J. Møller, P. Bacher. “Does using the app Watts influence electricity consumption? An investigation using linear mixed effect modelling”. DTU Technical Report 2019-xx.

Contents

Part I

Summary Report

CHAPTER 1

Introduction

1.1 Context and Motivations

1.1.1 Electricity Data

In the end of 2020, the smart-meter roll-out for electricity meters will finish in Denmark and since 2016 the supplier-centric DataHub [1] has been implemented. With this, service providers can get access to electricity consumption data with consumers permission and the potential of increasing awareness among residential electricity consumers is present. The data in the DataHub is usually in hourly resolution, available up-to one year in the past, and intended for billing purposes. The minimum requirement for the newly installed meters in Denmark is a temporal resolution of 15 minutes, but the utilities are only required to upload hourly measurements to the DataHub [2]. The smart-meter roll-out is also unfolding in the rest of Europa and several other EU countries are implementing similar datahubs as in Denmark [3, 4]. In EU there is no common agreement on the minimum temporal resolution, only a recommendation of 15 minutes [5]. Hence, it is unlikely that data with high resolution in the temporal domain will be available in the datahubs, without additional equipment to the smart-meters.

The variability of observed electricity consumption in similar residential buildings, can be very large. This is mainly due to the different behaviours of the occupants [6], which not only impact electricity consumption but the overall energy consumption of the building. Furthermore, electricity consumption for individual households is also likely to change over time, both due to individual life changing events, e.g., having children, renovations or appliance upgrades.

1.1.2 Long-term monitoring as feedback

There have been numerous studies on the effect of providing feedback to electricity consumers on their consumption. In [7] many studies are reviewed, and it is suggested that feedback to consumers on their electricity consumption leads to lower consumption. There are many different approaches in these studies and [7] suggest some likely features for successful feedback. A subset of these features is that the feedback is (list from Paper A):

- based on actual consumption
- given frequently
- given over a longer period.

A more recent study has also shown evidence that feedback can lead to reduction in consumption through an in-home display[8].

Motivated by this, the idea is to use the consumption data as a monitoring tool, such that consumers can act proactive on the development of their consumption. This is done by predicting the consumption for a period and then presenting the deviation between the actual consumption and the prediction. If the consumption is below the prediction, they get a positive message, and if it is significantly above, they get a warning, that their behaviour has changed leading to higher consumption. Instead of using the hourly data, predicting daily values are considered to reduce unnecessary complexity and still be able to pin point accurately when possible changes occur.

We have only seen few studies in the domain of mid to long-term forecasts for individual residents on daily electricity consumption using smart-meter data. This is confirmed by the review paper [9] on data-driven models for building energy consumption prediction, which concludes that research on long-term building energy consumption prediction and residential building energy consumption may require more attention. In [9] 63 studies were reviewed and only 3 were on residential electricity consumption on a daily level. [10] use Support Vector Regression for forecasting energy consumption in multifamily residences. In

[11] support vector machines are used on daily values for prediction one month ahead. In [12] multiple linear regression was studied for short-term forecasting in individual residences as input to a home energy management system. In these three papers models for daily prediction was studied, but not the performance on forecasts in long horizons. We also found two more studies. In [13] a bottom-up approach using clustering and data-mining of high-resolution data on appliance level is proposed. In [14] three different regression methods for hourly and daily consumption is studied. Unfortunately, no predictive performance is evaluated.

1.1.3 Disaggregated feedback

Another type of feedback is disaggregation. [15] reviewed several studies on disaggregated feedback of electricity consumption and suggests that providing this at appliance level leads to savings of electricity, hence looking into this domain is also of interest.

Disaggregation or non-intrusive load monitoring (NILM) [16] is the extraction of appliance level data from whole building/household energy signal using statistical approaches [15]. Disaggregation can be divided into two groups based on the temporal domain of the data, fine-grained disaggregation (time steps $\leq$ 1 minute) and coarse-grained disaggregation (time steps from 15 minutes to 1 hour) [17]. Fine-grained disaggregation will usually split a signal by detecting appliance signatures, either supervised or unsupervised [16]. Course-grained disaggregation divide the load into load correlating with outdoor temperature, continuous loads and time dependent loads [15, 17].

Research on disaggregation of electricity consumption has been explored for more than thirty years [17, 18, 19], but focus has previously mostly been on fine grained disaggregation. Lately more attention is given to coarse-grained disaggregation, e.g., [20, 21, 22, 23], these are all studies on disaggregation of hourly smart-meter data.

[20] use a supervised approach using data on appliance level for training. Via discriminative sparse coding, models for each appliance are trained and used to separate an aggregated signal. Supervised K-Nearest Neighbours is used in [21] to disaggregate in time steps of both 15 minutes and hourly. In [22] the whole-house electricity use was disaggregated into, in relation to external temperature, base load, activity load, cooling and heating season gradient. In [23] the consumption was disaggregated into a modelled heat load and other appliances based on regression models with weather data and response data from a household survey as input. [22] found three categories of loads, base load, activity load and temperature load without use of prior knowledge.

Since data on total household consumption, without prior knowledge on appliances or time-series on individual appliances, is building up in the DataHub more focus should be given to unsupervised coarse-grained disaggregation of activity and base load.

Looking for approaches on study and methods of disaggregation the field of occupant behaviour in buildings is of interest. In this field the focus is often on modelling presence in, or interaction with indoor climate. We have found that many studies are using Markov chains/processes in the description of the transition between presence, non-presence, movement between rooms and transitions between activities, e.g., [24, 25, 26, 27, 28, 29, 30]. This indicates that Markov chains/processes are highly useful for modelling occupant behaviour in a wide range of settings and has spurred us to look at methods to observe occupant behaviour and disaggregation in an indirect manner, e.g., Hidden Markov Models (HMMs).

1.1.4 Gamification of electricity consumption

Besides providing feedback to the consumer to help them save electricity, engaging or entertaining approaches could help the consumers become aware of their consumption patterns. Several studies on gamification and serious games within the domain of residential energy consumption are reviewed in [31]. From this review it is concluded that in general, there is a positive effect on reducing energy consumption, but also that there is a limited amount of empirical evidence suggesting more rigorous follow-up studies are needed.

1.1.5 Feedback channel for consumers

In late 2016 the app Watts [32] (Figure 1.1) was launched as a digital energy assistant for electricity customers at SEAS-NVE [33]. This app is the channel for providing feedback on electricity consumption to the consumers. The main feedback feature in Watts is the long-term prediction of the consumers consumption presented as the budget for the period. Deviation from the budget is presented as a bubble, coloured in accordance to the deviation so far in the period. Green if below the budget, yellow if above, and red if 30% above or more.

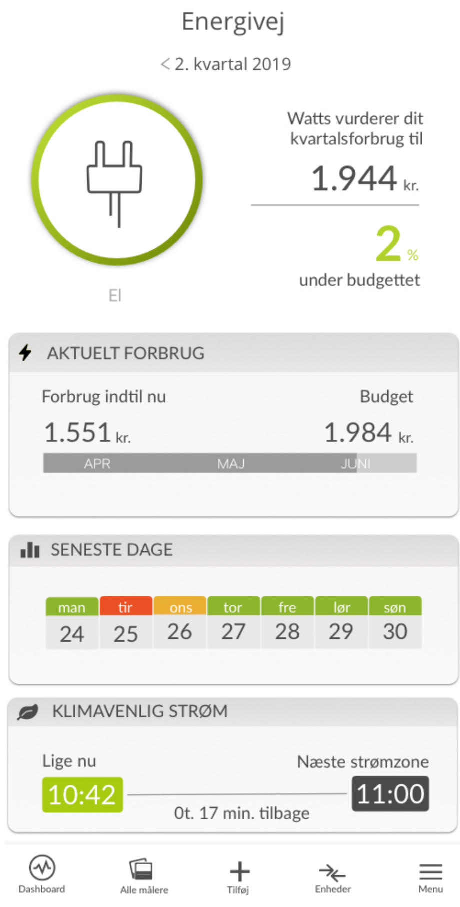


Figure 1.1: The Dashboard of the app Watts.

1.2 Thesis Objectives

The objective of this thesis is to develop methods for feedback/advice to electricity consumers based on their own consumption, investigate how users can be encouraged to interact with feedback and quantify the effect of the feedback, on the electricity consumption. To narrow the scope of this work, the feedback considered is a monitoring tool (long-term prediction), disaggregation of electricity consumption, gamification of electricity consumption, and methods for quantifying effect of feedback and advice. With this narrowing, four research questions are addressed in this thesis.

I. How can long-term prediction by applied in large-scale

Research have showed that feedback is needed over longer periods of time to be effective, hence tools for monitoring electricity consumption are needed. By making long-term predictions of the individual consumers consumption, the consumer can monitor her consumption in relation to the prediction and deviation from the prediction can warn the consumer of changes in consumption patterns. There have previous been little attention to development of long-term prediction models for individual households, there is a great diversity in consumption patterns and people change consumption patterns over time. Hence, adaptive models for medium to long-term predictions, robust to the many patterns, will be developed.

II. Can hourly metering data be disaggregated into meaningful feedback?

Several studies have showed that disaggregated feedback of electricity consumption can lead to saving electricity, but there is a lack of research on coarse grained disaggregation addressing the temporal resolution available in the datahubs being implemented around Europa, in particular the activity load and base load consumption. Models addressing this will be developed.

III. Can gamification of electricity consumption increase awareness and trigger changes in behaviour

It is of interest to investigate other means of engaging consumers to increase awareness on their electricity consumption, here gamification of electricity consumption could be an approach. Hence, approaches of gamification of electricity consumption will be considered.

IV. Does the provided feedback influence the electricity consumption?

It is a key objective to quantify if there is an effect of the developed feedback and answer the question; is Watts an effective monitoring tool? Hence, this will be investigated.

1.3 Thesis Contributions

The modelling framework for long-term prediction of electricity consumption, using adaptive recursive least squares developed in Paper A is used in the App Watts (currently for 45.000 smart-meters) with extensions described in Paper B were solutions to model deficiencies found in the large scale test of the modelling framework in Paper A is addressed.

In Papers C and D coarse grained disaggregation is addressed. Paper C addresses the disaggregation from the field of occupant behaviour modelling and showed potential for probabilistic short-term forecasting and simulation. In Paper D coarse grained disaggregation with focus on appliance activity is addressed. The models from Papers C and D are yet to be implemented in such a way that the disaggregated feedback can be presented to the users.

In Paper E a prototype game was developed and tested on four Watts users, using consumers own consumption as part of the game play. Although it was difficult for the players to understand the game in the current state, it gave rise to ideas on how to engage consumers through gamification.

We address the effect of the app Watts, on electricity consumption in Paper F, where a significant effect was found.

1.4 Thesis Structure

This thesis is structured as follows. Part I is a summary report outlining the main contributions of this thesis. Chapter 2 Describes the developed methodology for medium to long-term prediction of residential electricity consumption for large scale application. In Chapter 3 the methodologies for disaggregation of hourly electricity consumption are described. An approach for gamified interaction with electricity consumption is studied and the effect of using the app providing feedback is investigated in Chapter 4 . Chapter 5 provides conclusions and perspectives.

Part II consists of the publications that contribute to this thesis.

Paper A has been submitted for consideration to *Applied Energy*. A model framework for medium to long-term prediction of electricity consumption is developed using adaptive Recursive Least Squares and applied in large-scale.

Paper B is a technical report, pending to be published by the *DTU Library*. Extension to the Recursive Least Squares model framework are introduced and examples of their benefits are presented.

Paper C is a journal article published in *Sustainable Cities and Society*. By using Hidden Markov Models, electricity consumption is disaggregated into states describing the occupants' presence in the household.

Paper D has been submitted for consideration to *Applied Energy*. Using Continuous Time Hidden Markov Models, electricity consumption is disaggregated into states describing possible appliance related activities.

Paper E is a technical report, pending to be published by the *DTU Library*. An approach for gamification of electricity consumption and the development of a game is presented. The game is tested on app users and evaluated.

Paper F is a technical report, pending to be published by the *DTU Library*. Using linear mixed effect models the effect on electricity consumption of using the app Watts is investigated

Chapter 2

Medium to long-term prediction of electricity consumption

In this chapter the adaptive Recursive Least Squares with forgetting (from now on denoted RLS) modelling scheme is introduced in Section 2.1. In Section 2.2 the scheme is tested for large-scale application. From this the performance is discussed and model deficiencies are highlighted. In Section 2.3 the deficiencies are addressed and extensions to the modelling scheme are presented. With these extensions the impact on the large-scale application is presented. Lastly an example of heat load disaggregation based on the RLS modelling scheme is presented in Section 2.4

2.1 Introducing the RLS model procedure

In this section the modelling scheme is presented. The purpose of the modelling scheme is to find and tune a suitable model for each individual household. Based on a exploratory data analysis (Section A.2) the following output and input variables was chosen as possible explanatory variables (adapted from Paper A):

- Output variable:
  - Electricity consumption.
- Input variables:
  - Temperature difference between a temperature threshold and the daily average ambient temperature (Equation (A.1)).
  - Low pass filtered temperature difference to model the envelope of the house (Equation (A.2)).
  - Length of the night given location of the household.
  - Fourier series for describing the weekly variation (Equation (A.11)).

The modelling scheme is divided in to two stages. In the first stage, a model structure for each household, which input variables to enter the model, is found using a forward selection approach. In the second stage, optimization of the RLS for each individual horizon, by choosing the harmonics in Fourier series describing the weekly pattern, is carried out.

2.1.1 Stage 1: Model structure

For each household a linear regression model is fitted using forward selection with the Bayesian Information Criteria (BIC) as the model fit criterion. The BIC is a measure of the relative fit of a statistical model to a given set of data. The forward selection works by starting with an intercept model and in each step adding the variable that will improve the BIC value the most, until no more improvement is possible. If temperature dependence is found during forward selection, then either the temperature difference is chosen or the filtered temperature difference is chosen, not both.

When the best model has been selected the estimated parameters are checked. Except for the intercept parameter, they must all be positive in order to be in line with the physical effects, e.g., a higher temperature difference should result in a higher consumption from a physical point of view. Hence negative correlations to the selected explanatory variables does not make sense and are therefore removed from the model for that household (see Sections A.2.1.1 and A.2.1.2 for descriptions). If temperature difference or low pass filtered temperature difference is chosen, an optimal temperature threshold is found for the individual household.

The first stage is presented as pseudo code in Algorithm 1. See the nomenclature in Paper A for notation.

Algorithm 1 Choosing explanatory variables using forward selection [Paper A]

Input:
$\mathbf{y}$, training data model output of length n.
$\mathbf{X}$, full training data input, all possible variables.
Output: $\mathbf{X}_{\text{RLS}}$, the reduced training data input for RLS.

Set $T_{\text{threshold}} = 22$
$\mathbf{T}_d \vee \mathbf{T}_f$ by Equation (A.1) and (A.2)
$\hat{\boldsymbol{\theta}}, \mathbf{X}_{\text{Reduced}} := \mathit{Forward.selection}(\mathbf{y}, \mathbf{X})$
if $\mathbf{T}_d \vee \mathbf{T}_f \in \mathbf{X}_{\text{Reduced}}$ **then**
$\quad \arg\min_{T_{\text{threshold}}}(\mathit{BIC.LinReg}(\mathbf{y}, \mathbf{X}_{\text{Reduced}}))$ for $T_{\text{threshold}} \in [5, 22]$,
end if

$A := \{1, \dots, \dim(\hat{\boldsymbol{\theta}})\}$, A is the set of indices of the parameter vector.
$B := \{i \in A | \hat{\theta}_i > 0 \vee i = 1\}$, B is a subset of A with intercept ($i = 1$), and where $\hat{\boldsymbol{\theta}}$ have positive values.
$\mathbf{X}_{\text{RLS}} := \mathbf{X}_{\text{Reduced},*,B}$, The input matrix is reduced to the columns with indices in B.

return $\mathbf{X}_{\text{RLS}}$

2.1.2 Stage 2: Recursive Least Squares with forgetting

To introduce Recursive Least Squares with forgetting the description from Paper A is presented below:

RLS with forgetting is a parametric model where the implementation is based on [34]. The RLS algorithm is presented for horizon k

Update step:

$$\mathbf{R}_{t-k} = \lambda \mathbf{R}_{t-k-1} + \mathbf{x}_{t-k}\mathbf{x}_{t-k}^T, \tag{2.1}$$

$$\hat{\boldsymbol{\theta}}_{t-k} = \hat{\boldsymbol{\theta}}_{t-k-1} + \mathbf{R}_{t-k}^{-1}\mathbf{x}_{t-k}(y_{t-k} - \mathbf{x}_{t-k}^T\hat{\boldsymbol{\theta}}_{t-k-1}). \tag{2.2}$$

Prediction step

$$\hat{y}_{t|t-k} = \mathbf{x}_t^T\hat{\boldsymbol{\theta}}_{t-k}. \tag{2.3}$$

Here $\hat{\boldsymbol{\theta}}_t$ *is the estimated parameter vector,* y_t *is the observation,* $\hat{y}_{t|t-k}$ *is the prediction,* $\mathbf{x}_t$ *is the vector of observed input,* $\mathbf{R}_t$ *is the inverse sample covariance matrix up to a constant and* λ *is the forgetting factor. One model* *for each horizon k is fitted, hence for a quarterly prediction we need approximately 92 model fits for each household depending on quarter and year [Paper A].*

In this stage the RLS is fitted optimising using root mean squared error (Equation A.10) as the objective function, choosing between the number of Fourier series to describe the weekly pattern.

The second stage is presented as pseudo code in Algorithm 2. See the nomenclature in Paper A for notation.

Algorithm 2 RLS modelfit [Paper A]

Input:
$\mathbf{y}$, training data model output of length n.
$\mathbf{X}_{\text{RLS}}$, training data input.
k, the prediction horizon in number of days.
p, number of sine, cosine pairs of the Fourier series input.
$\mathbf{X}_p$, matrix of p pairs of Fourier series training data.
Output: *ModelList*, list of model fits of length k.
for $i = 1$ **to** k **do**
  for $j = 0$ **to** p **do**
    $\mathbf{X} := [\mathbf{X}_{\text{RLS}}\mathbf{X}_p]$, Fourier series input is combined with the input found in Stage 1.
    $\hat{\boldsymbol{\theta}}_n^j, \hat{\mathbf{y}}^j := RLS(\mathbf{y}, \mathbf{X})$, The RLS is fitted for each added pair of Fourier series input.
  end for
  $q := \arg\min_j (RMSE_i(\mathbf{y}, \hat{\mathbf{y}}^j))$, The fit with smallest RMSE is found.
  ${}_i\hat{\boldsymbol{\theta}}_n := \hat{\boldsymbol{\theta}}_n^q$
end for
return *ModelList* $:= \{{}_1\hat{\boldsymbol{\theta}}_n, \cdots, {}_k\hat{\boldsymbol{\theta}}_n\}$

The *ModelList* is then used with Equation (2.3) and prediction of the input to calculate the prediction of the model output.

2.2 Large-scale application of the modelling framework

The presented modelling scheme was applied on almost 22000 times-series of household electricity consumption. For each household data from entire 2014 and 2015 was available. For each quarter of 2015 a model using data one year back from the beginning of the quarter was selected and fitted for each household, hence quarterly predictions (90-92 days ahead) was obtained for all four quarters.

For some of the households the models gave negative predictions, In Table 2.1 the number for each quarter is shown, these are removed in the evaluation. Furthermore, households with long absence in the benchmark or prediction periods are removed, since comparing with such periods is not of interest. Counts of the different model structures from stage 1 of the modelling scheme are shown in Table 2.1.

Table 2.1: Counts of the selected model structures in each of the four predicted quarters. The structure describes which input variables were chosen in the forward selection procedure. A 1 denotes the intercept, LN denotes length of the night, T_{d} is the temperature difference and T_{f} is the filtered temperature difference. The Fourier input is not shown. [Paper A]

Quarter	1	2	3	4
Input structure				
(1)	1597	2096	1809	1712
$(1, LN)$	5644	5685	5982	5735
Subtotal	(7241)	(7781)	(7791)	(7447)
$(1, T_{\mathrm{f}})$	4218	3343	3139	2991
$(1, T_{\mathrm{f}}, LN)$	7707	7612	6963	7881
$(1, T_{\mathrm{d}})$	677	703	771	756
$(1, T_{\mathrm{d}}, LN)$	1529	1787	2120	2344
Subtotal	(14131)	(13445)	(12993)	(13972)
Total	21372	21226	20784	21419
Reported electrical heating	9009	8905	8642	9013
Removed ($\hat{y}_i < 0$)	436	618	1061	424

2.2.1 Benchmark and performance measures

In order to understand how the predictive performance of the model it must be tested on a large set of households. Two performance measures are used to compare the predictions from the model vs. the benchmark, as well as between the groups of different model structures. The benchmark is the average consumption from the same period (i.e., same quarter one year ago, see Equation (A.13)). This benchmark is selected because it is the best naive predictor, since it is adapted to each household and easy to implement.

The first measure is a relative measure of prediction bias. The Relative Cumulated Error (RCE) (Equation (A.14)). With a relative measure it is possible to compare the performance of the model to the benchmark and directly between the households. Hence, it is easy to make groupings based on model structures that have entered the model and compare them in order to identify model deficiencies and possible improvements.

To validate the accuracy on the daily prediction the Sum of Squared Errors Ratio (SSER) (Equation (A.15)) between the prediction SSE and the benchmark SSE. This is also a relative measure that can be directly compared between households. If the SSER is less than one, then the RLS model is performing better than the benchmark.

2.2.2 Predictive performance

The empirical distributions of the RCE calculated for all the households are shown as box-plots in Figure 2.1, where the model predictions and benchmark RCE are compared. From this it is clearly seen that there very is little difference between the distributions for the benchmark and the model predictions. Also, they seem centred around zero, hence no noticeable bias is found.

The same type of box-plots are generated for the model structures listed in Table 2.1. These box-plots are shown In Figure 2.2, again the model predictions and benchmark RCE are compared for each model structure and quarter. The general picture is that some variation between the distributions now can be seen and in overall it seems that the benchmark and the model predictions share the same tendencies, i. e., they are under- and over-predicting in the same quarters. For the temperature dependent model structures $(1, T_{\mathrm{f}}, LN)$ and $(1, T_{\mathrm{d}}, LN)$ (the two plots to the right) the biggest difference is seen in Q3, which is the summer quarter in Denmark, but no big difference is expected, since cooling is rare. From this it shows that the model predictions are in general not very

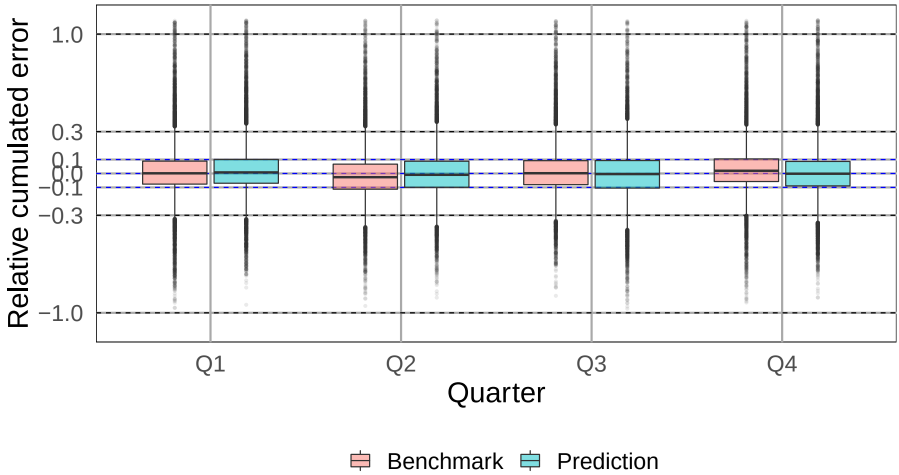


Figure 2.1: The empirical distribution of the RCE for all households for each predicted quarter, both for the benchmark and the model predictions [Paper A].

different, in terms of bias, from the benchmark.

It was investigated if the impact of temperature, should give rise to differences in bias between the benchmark and the model. It was found most likely that there was not a significant difference in temperatures between the two years, but only with the two years to compare it was not found possible to conclude upon.

2.2.3 Daily predictions

To analyse the prediction performance on a daily level the SSER (Equation (A.15)) is calculated for each household in each quarter and grouped on model structure. Box-plots of empirical distribution of these are shown in Figure 2.3. If the SSER is smaller than one, then the model performs better than the benchmark.

For the intercept model and the model with only length of the night as input (the two leftmost plots) the median is slightly less than one, indicating no significant differences in prediction accuracy between the benchmark and the model predictions.

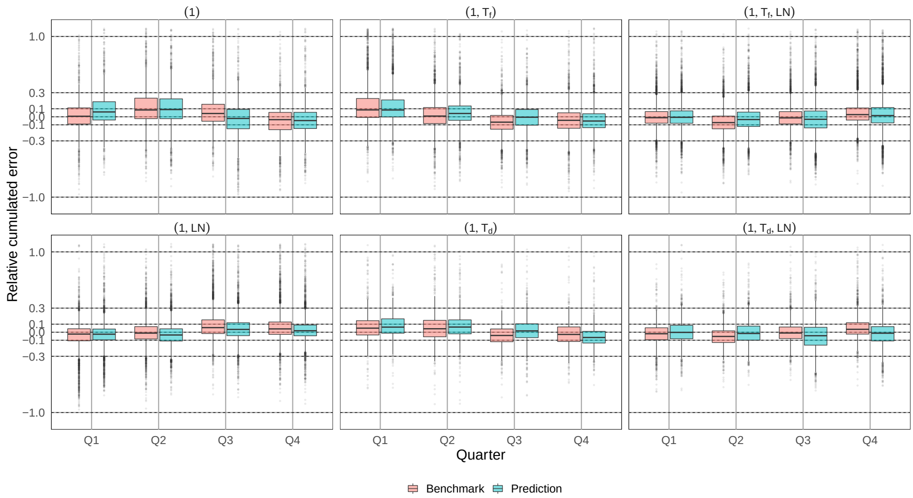


Figure 2.2: The empirical distribution of the RCE of the model prediction compared to the RCE of the benchmark. Each plot shows the distribution of the RCE given the selected model structure indicated above the plot [Paper A].

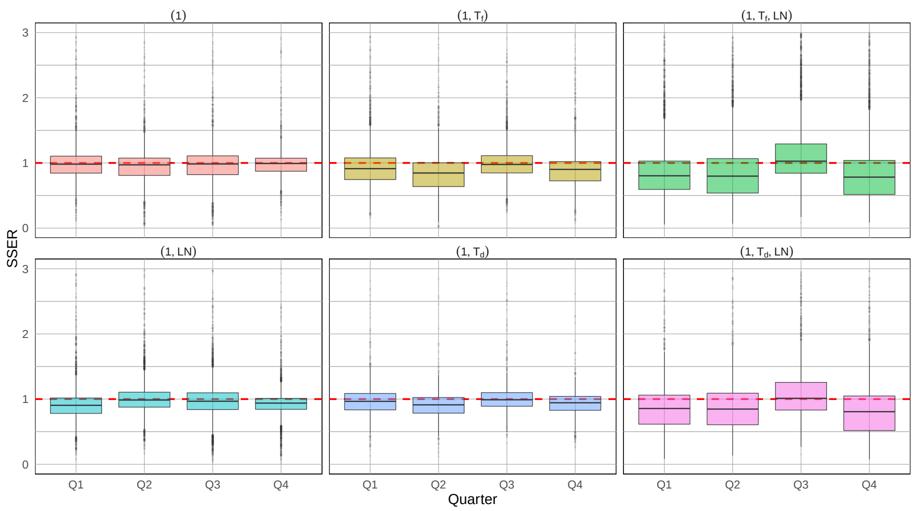


Figure 2.3: The empirical distribution of the SSER, given model structure and quarter, Each plot shows the distribution of the SSER given the selected model structure indicated above the plot [Paper A].

For the temperature dependent model structures (the two middle plots) the median is below one for all quarters, hence the model predicts in general more accurate than the benchmark of daily levels.

Finally, for the temperature and night length dependent model structures (the rightmost plots) a clear general improvement of model over the benchmark is found, except for Q3, which is meaningful, since this is the summer quarter and thus the temperature should not impact the predictions.

2.2.4 Deficiencies

To highlight model deficiencies an analysis on prediction deviation from the observed consumption was carried out in Paper A. Four examples highlighting the most common are presented (list adapted from Paper A):

- Figure 2.4a: It can be seen that a change in consumption close to the prediction period, around May 2015, is not caught by the model.
- Figure 2.4b: A deficiency of the model can be seen, where the impact from temperature explode on longer horizons. This can be categorized as a kind of Type I error, since data by chance fell out such that the impact was highly over estimated.
- Figure 2.4c: Another example where the residents are absent in periods. The longer horizons are affected by this leading to a decreasing level during the prediction period.
- Figure 2.4d: The intercept is very low or negative and the model has not adapted to the non-heating season, this is often also the case for negative prediction although not in this particular case.

To summarise the findings from this analysis, the most common deficiencies found analysing the performance of the large-scale application is:

- Slow adaptation
- Negative/low intercept in non-heating season due to high temperature impact
- Exploding impact of temperature in the long horizons.

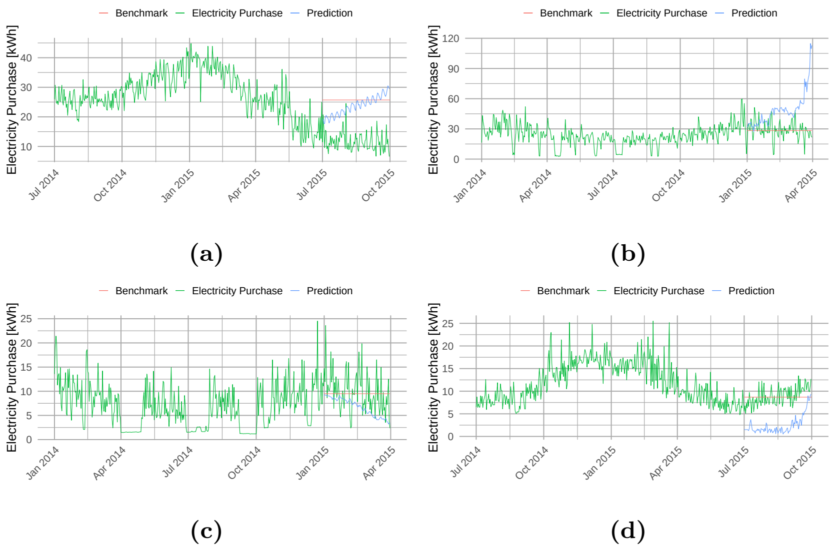


Figure 2.4: Examples where the model scheme is deficient [adapted from Paper A].

2.3 Extensions to the modelling scheme

The model deficiencies found are addressed in Paper B. First, self-tuning control of the RLS (STCRLS) [35] is employed for faster adaptation, see Equations (B.5) to (B.15). Here upper and lower bounds on the co-variance matrix, with means of tuning this if the bounds are reached and online tuning of the forgetting factor are implemented. Second, to avoid negative prediction values or very low predictions in the Q3 for temperature dependent household a Box-Cox transformation (Equation (B.16)) scheme is implemented optimising the transformation coefficient such that the transformed data resembles a normal distribution the best. It should be noted that when transforming the predictions back it is no longer a prediction of the mean value but a prediction of the median. Finally, using parameter coefficients from short horizon models to predict long horizon is tested.

2.3.1 RLS with Self-tuning control

To illustrate the possible improvement using self-tuning control an example is presented in Figure 2.5. There are several jumps in the consumption over time where the ordinary RLS have very slow adaptation to these changes. It is clear that the STCRLS quickly adapts.

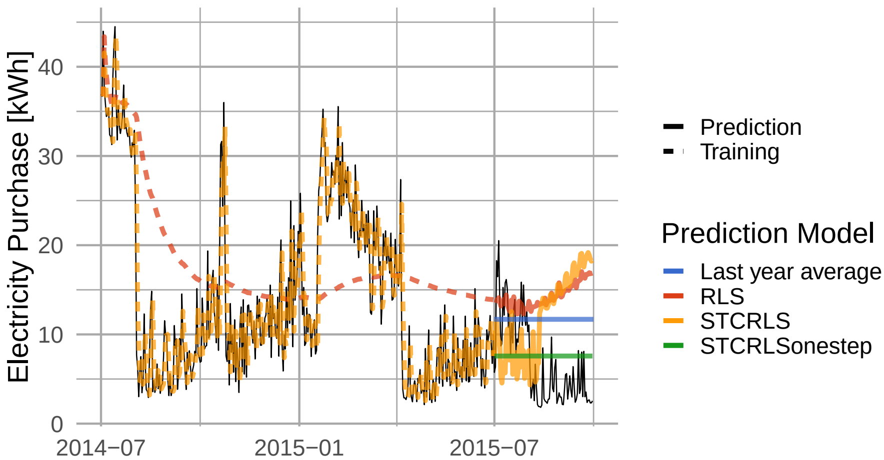


Figure 2.5: Adaptive behaviour for the 1 step horizon model of the ordinary and self-tuning RLS and predictions [Paper B].

This example also illustrates the benefit of using short horizon parameter coefficients for predicting each day in the period, since the long horizons of the STCRLS is still not relying on most recent data. The performance measures are summarised in Table 2.2.

Table 2.2: Relative cumulated error (RCE) and sum of squared error (SSE) for each model type shown in Figure 2.5 [Paper B].

Model type	RCE	SSE
RLS	1.09	7738.9
STCRLS	0.76	8720.7
STCRLSonestep	0.11	2090.5
Last year average	0.73	4213.6

2.3.2 Box-Cox transformation of data

To illustrate the possible improvement using Box-Cox transformation of the data an example is presented in Figure 2.6. The example shows a the ordinary

RLS yielding negative predictions. The issue for the RLS is that the intercept has become negative during the training for the short horizons, but for the long horizons the intercept is positive as if it was in the middle of winter and we get very high predictions. The transformed model fit is named RLSBC. Here great improvement on the short horizons are observed, but there is still an issue with high values of intercept in the long horizons and the SSE is high (Table 2.3). Using the parameter coefficients for $k = 1$ of the RLSBC shown in Figure 2.6 as RLSBConsetep the explosion is avoided, but the RCE shows an underestimation only prediction the lower values of the day to day variation.

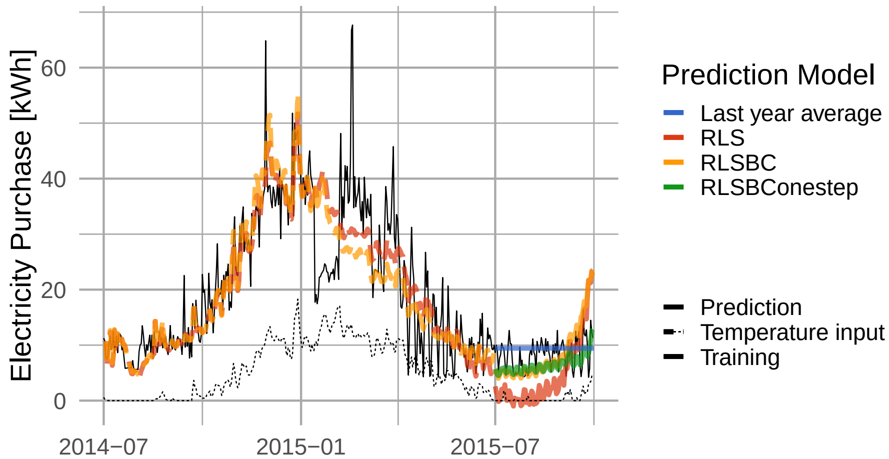


Figure 2.6: Adaptive behaviour for the 1 step horizon model of the ordinary with raw and Box-Cox transformed data and predictions [Paper B].

Table 2.3: Relative cumulated error (RCE) and sum of squared error (SSE) for each model type shown in Figure 2.6 [Paper B].

Model type	RCE	SSE
RLS	NA	4587.8
RLSBC	-0.01	2192.2
RLSBConestep	-0.23	886.6
Last year average	0.12	778.8

2.3.3 Performance of extensions

Based on the extension from Paper B, the modelling scheme is extended to eight model types with same procedure for choosing model structure.

Table 2.4: The different model types available with the extension the modelling scheme.

Model type	Description
RLS	Raw data fitted with ordinary RLS for all horizons.
RLSonestep	Raw data fitted with ordinary RLS only for horizon $k = 1$.
STCRLS	Raw data fitted with self-tuning RLS for all horizons.
STCRLSonestep	Raw data fitted with selftuning RLS only for horizon $k = 1$.
RLSBC	Box-Cox transformed data fitted with ordinary RLS for all horizons.
RLSBConestep	Box-Cox transformed data fitted with ordinary RLS only for horizon $k = 1$.
STCRLSBC	Box-Cox transformed data fitted with self-tuning RLS for all horizons.
STCRLSBConestep	Box-Cox transformed data fitted with self-tuning RLS only for horizon $k = 1$.

For all these model types predictions for each household and quarter was carried out. For each model type the empirical distributions of the RCE are calculated for all the households and shown as box-plots in Figure 2.7, where the model predictions and benchmark RCE are compared. From this there is no clear single model type standing out. There are little difference between the distributions for the benchmark and the model predictions. The Box-Cox model types seem to be biased toward under estimation which is expected due to predicting the median instead of prediction the mean.

Since no one model type stood out on performance and we have seen that in certain situations the extensions are working well, for each household and quarter the model type with the best RCE was chosen. This is to see how performance would be if we had perfect information. The empirical distributions of these RCE are shown in Figure 2.8 and the empirical distributions of these SSER are shown in Figure 2.9.

From Figures 2.8 and 2.9 the distributions around zero is much smaller for the model RCE and all the mean values of the SSER is now below one. The potential performance of the modelling scheme is highlighted and indicates that great improvement of the modelling scheme is possible if methods for choosing an optimal model type is developed.

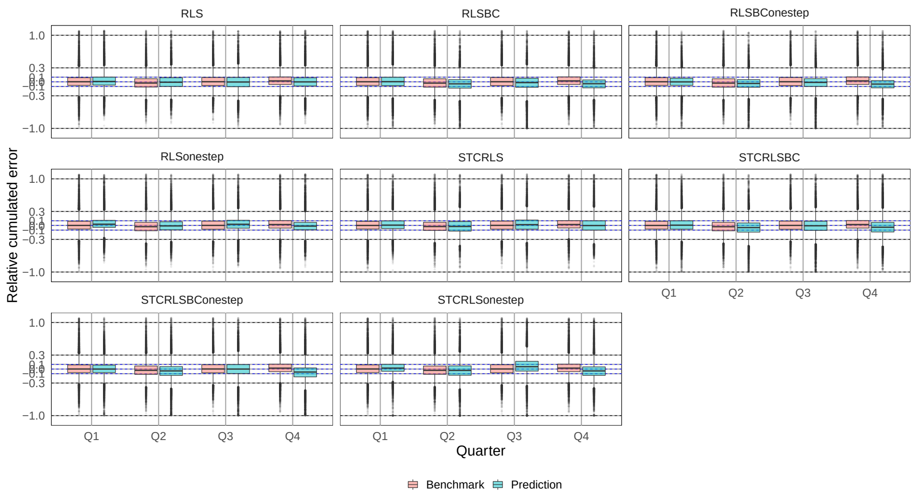


Figure 2.7: The empirical distribution of the RCE for all households for each predicted quarter, both for the benchmark and the model predictions for each model type.

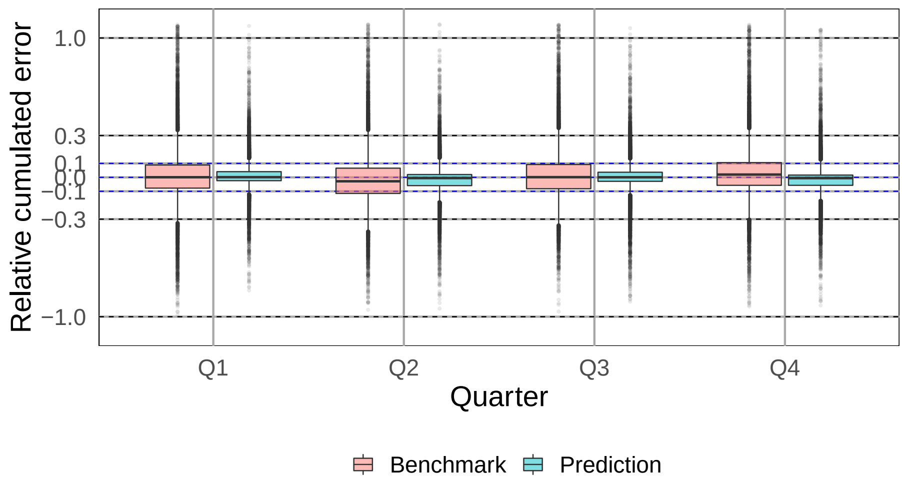


Figure 2.8: The empirical distribution of the RCE for all households for each predicted quarter, both for the benchmark and model predictions, for the model type with best RCE for each household.

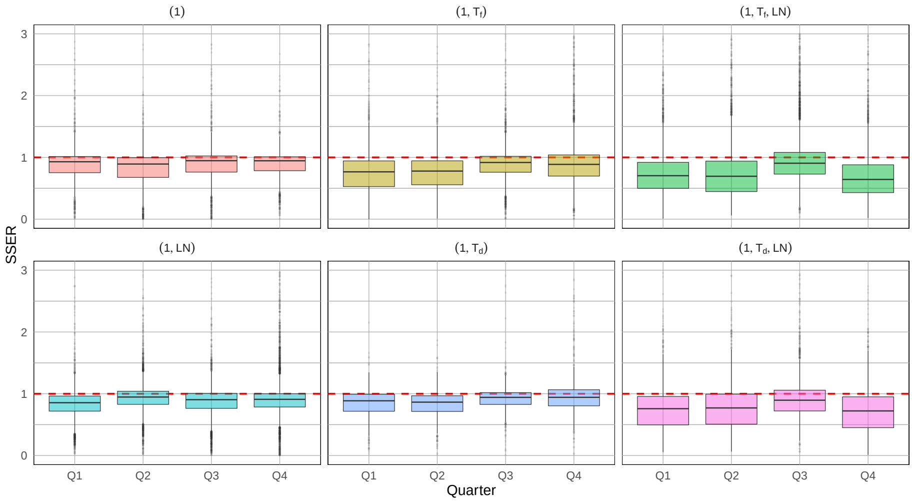


Figure 2.9: The empirical distribution of the SSER for the model type with best RCE for each household, given model structure and quarter, Each plot shows the distribution of the SSER given the selected model structure indicated above the plot.

2.4 Temperature disaggregation

As an example on how the RLS model scheme can be utilised in retrospective, a disaggregation of the heat-load for a temperature dependent household in the beginning of winter is carried out and shown in Figure 2.10.

In Figure 2.10 a clear stable activity load is seen, this is the daily expected consumption without heating, in this example no clear weekly variation was found to enter the model, but some variation is seen in the deviation load which are the residuals of total predicted load and the actual load. The heat-load corresponds well with the temperature input, only in the end of November some greater deviation is observed. First a period of under estimated consumption and then a short period of over estimation, this could possibly be due to absence in a short period where the ambient temperature dropped cooling the house below comfort level, thus the heat is turned up to return to the comfort level when the residents returned. This is of course speculations, but the information could help pin point for the residents when unusual consumption occurs end be dealt with if necessary.

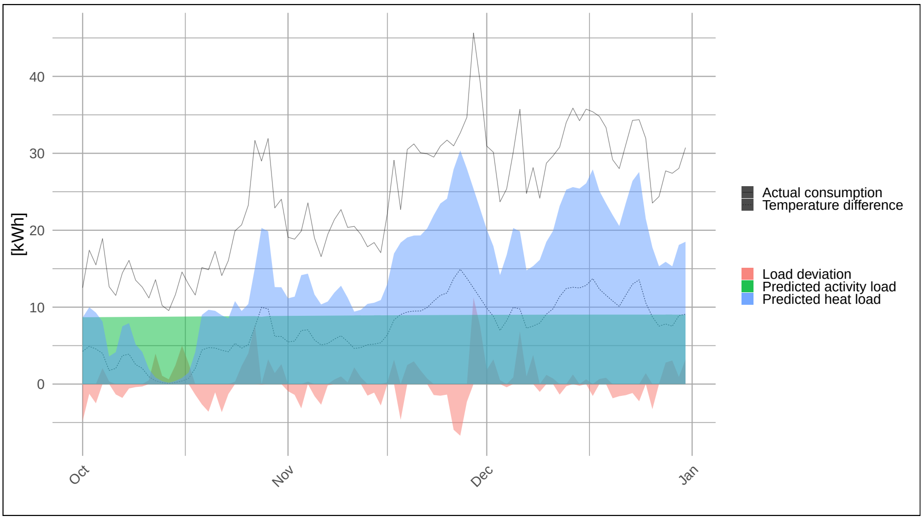


Figure 2.10: Retrospective disaggregation using the parameter coefficients updated to the last day in the period. The temperature threshold was found to be 13.8°C for this particular households.

CHAPTER 3

Disaggregation of electricity consumption

This chapter addresses course grained disaggregation of smart-meter data on electricity consumption in hourly temporal resolution. In Section 3.1 Both Discrete-Time (DT), Continuous-Time (CT) Hidden Markov Models (HMM) and the key inference tool global decoding are introduced. In Section 3.2 DT-HMMs are applied on electricity consumption of 14 apartments from Catalonia, Spain. A CT-HMM is applied on a single apartment from Sweden in Section 3.3. Finally, In Section 3.4 the main differences between DT-HMM and CT-HMM are summarised.

3.1 Introducing Hidden Markov Models and global decoding

3.1.1 Discrete Time Hidden Markov Models

DT-HMMs consists of an independent mixture distribution and a Markov chain. A Discrete Markov chain is A sequence of discrete random variables $\{C_t : t \in \mathbb{N}\}$

if, for all $t \in \mathbb{N}$, the Markov property is satisfied, i.e.

$$Pr(C_{t+1}|C_t, ..., C_1) = Pr(C_{t+1}|C_t). \tag{3.1}$$

The probabilities for changing between states can be collected in the transition probability matrix $\mathbf{\Gamma}$,

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix}, \tag{3.2}$$

for m states where each row represent the probability of staying in the given state (γ_{ii}) or changing to another state (γ_{ij} for $i \neq j$). The Markov chain is called homogeneous if the transition probabilities are not dependent on time, otherwise it is called inhomogeneous.

For each state an underlying probability distribution describes the observed data for the given state (Figure 3.1) and the transition probabilities describes the mixture.

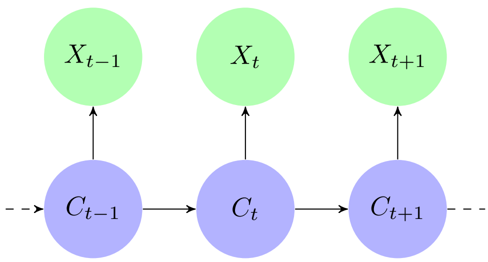


Figure 3.1: Directed graph of a Hidden Markov Model [Paper C].

For the transition probability matrix $\mathbf{\Gamma}$, there are m^2 entries, but only $m(m-1)$ free parameters due to the row-sum constraint. In total we end up with $m(m-1) + mp$ parameters where p is the number of parameters in the individual state-dependent distributions. This is a quadratic growth of parameters with m.

Implementing inhomogeneous HMMs and adding co-variates to the transition matrix increases the number of parameters. For more in-depth description see Section C.2.

3.1.2 Continuous Time Hidden Markov Models

For continuous-time Markov chains, transitions may occur at all times. In the continuous case the quantities of interest become the transition intensities which are the derivatives of the transition probabilities when $t \to 0$. Assuming it is only possible to change to neighbouring states in a short time interval the transition intensities can be described as

$$\boldsymbol{Q} = \begin{pmatrix} -q_1 & (1-w_1)q_1 & 0 & \cdots & 0 & w_1 q_1 \\ w_2 q_2 & -q_2 & (1-w_2)q_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ (1-w_m)q_m & 0 & 0 & \cdots & w_m q_m & -q_m \end{pmatrix}. \tag{3.3}$$

It is seen that now we end up with $2m + mp$ parameters, i.e., growing linearly with m. With this set up, change to non-neighbouring states can occur but the transition must go through a neighbouring state.

The sojourn time is the expected time a continuous-time Markov chain will stay in a given state. When the process enters state i, it stays there according to the sojourn time, before it jumps to another state $j \neq i$ with probability q_{ij}/q_i. For further details on CT-HMM see Section D.2.

3.1.3 Global decoding

Global decoding is the key inference tool, used for retrospective analysis. For a HMM and observations, the most likely state for a given time t can be determined by looking at the probability of observing this realisation for each state and choosing the one with the highest probability. This is called local decoding. Global decoding is the determination of the most likely sequence of states conditioned on all the observations, i.e., now the probability of transition from the previous most likely state is also considered for $t \in \{1, ..., T\}$ (See Section D.2.6 for further detail).

Given a most likely sequence of states and grouping these in accordance to time of day, a distribution of observed states for each hour of the day is obtained. The proportion of observations of a state for a given hour of the day is then an indication of the probability of observing this state in the given hour. Combining the observed proportions for all hours of the day we obtain a daily proportion profile for each state. With these profiles we get an indication of the most likely state for a given hour of day without fitting an inhomogeneous HMM. Using the daily proportion profile in combination with the knowledge of the HMM model parameters we will be able to infer knowledge of the states in relation to the underlying activities leading to the observed electricity consumption.

3.2 Application of Discrete-Time Hidden Markov Models

In Paper C Discrete-Time Hidden Markov Models is applied on hourly electricity consumption from 14 apartments. Gamma distributions were found most suitable for the state distribution and HMM with 3-4 states is found for each apartment. For estimation procedure and model selection see Paper C.

3.2.1 Estimated Parameters

To get an overview of the estimated parameter a summery for Apartment 2 is presented in Table 3.1. Here γ_{ij} is the probability for transition from state i to j, δ is the stationary distribution - the probability for being in a given state if no knowledge of prior state is known. k and θ are the independent state distribution parameters for gamma distributions.

Table 3.1: Estimated parameters for the three state HMM and calculated stationary distribution, mean and variance, Apartment 2 [Paper C].

State	k	θ	γ_{i1}	γ_{i2}	γ_{i3}	δ	Mean	Variance
1	7.74	0.012	0.85	0.14	0.01	0.52	0.09	0.001
2	7.30	0.040	0.21	0.73	0.06	0.38	0.29	0.012
3	5.14	0.205	0.00	0.30	0.70	0.10	1.05	0.216

From the mean and variance the consumption level is seen for each state and a first interpretation of the states could be (list from Paper C):

1. Low consumption
2. Medium consumption
3. High consumption

for states 1, 2, and 3, respectively.

3.2.2 Global decoding

Finding the most likely sequence of states using global decoding, inference on possible dependencies of the states can be found. In Figure 3.2 a clear diurnal dependence is observed suggesting a inhomogeneous HMM could be applied. Furthermore, the high load state seems to be dependent on temperature suggesting an air-conditioner is used when it is hot.

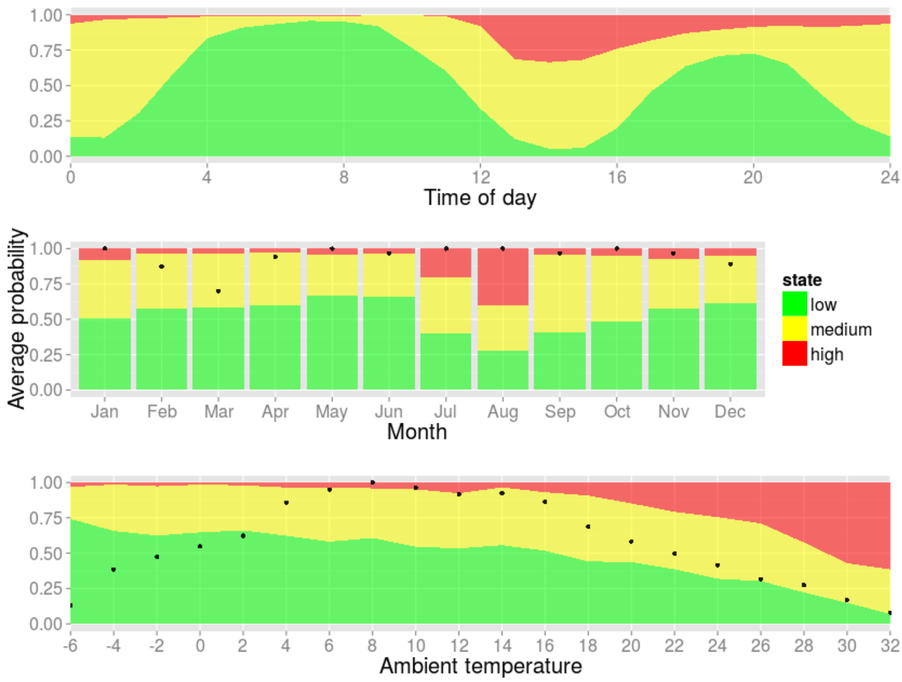


Figure 3.2: Average probability profiles of being in given state, dependent on time of day, time of year and ambient temperature, Apartment 2. The black dots denote the relative amount of observations for each month or temperature interval [Paper C].

By comparing the states with counts of water use given time of day in Figure 3.3 a new set of descriptions is found in relation to presence in the apartment. except from the morning hours there is no or little water use in state 1 indicating absence or inactivity in the apartment. The new set of descriptions is given below (list from Paper C):

1. absent or asleep
2. home, medium consumption
3. home, high consumption

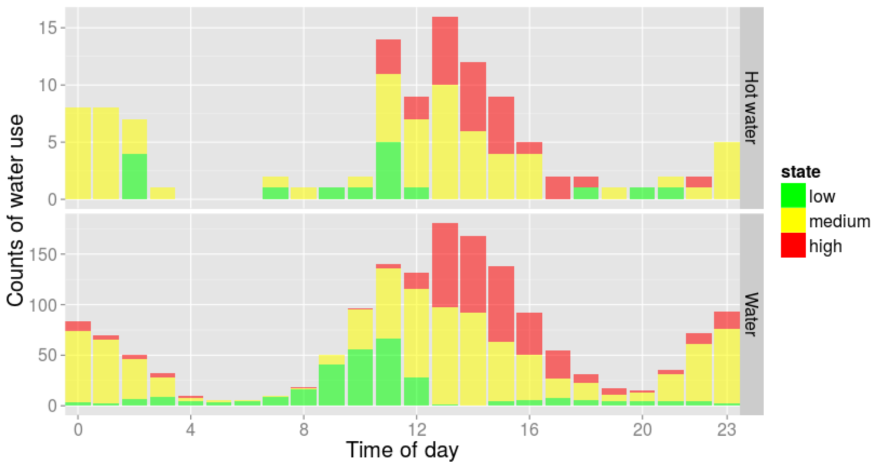


Figure 3.3: Counts of water and hot-water use given state and time of day, Apartment 2 [Paper C].

For the rest of the apartments daily dependence of the states was found and the same state descriptions.

3.2.3 Classification of daily patterns

Comparing the daily profiles of observed states similar patterns was observed between apartments and four distinct profiles were observed, these are presented in Figure 3.4 and described in Table 3.4.

The descriptions corresponded well to an occupant survey on number of resident and how many hours the occupants are absent on a weekday. With these observations' indications of possible random or fixed effects, that could enter a population model, where several apartments can be collected on one model with both common parameters and apartment specific parameters.

Table 3.2: Apartments classified based on the average probability profile given time of day [Paper C].

Class	Apartments
afternoon/evening absence	2
equal probability for being home or absent	1, 5, 7, 26 and 35
mostly at home	3, 18, 29 and 44
mostly absent	12, 15, 25 and 30

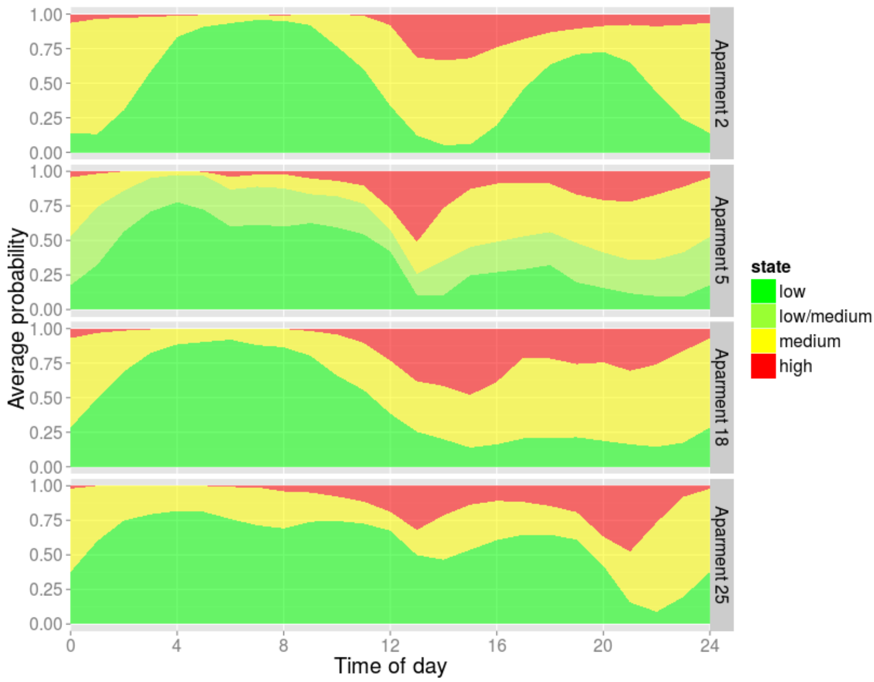


Figure 3.4: Observed distinct profiles. The low/medium state of Apartment 5 is considered low based on the mean value of the state-dependent distribution [Paper C].

To further investigate if the homogeneous HMMs have common parameters box-plots is produced for the three-state models in Figures 3.5 and 3.6.

For the state dependent distribution parameters (Figure 3.5) there is high variation for the k parameter in the low and medium state but not in the high state. For θ it is opposite. this indicates that k for the high state and θ for the low and medium state could enter as common parameters in a population model.

For the transition probabilities in Figure 3.6, the low state parameters could be common in a population model, it is not obvious for the medium and high state, except for the transition probability of going from high to low.

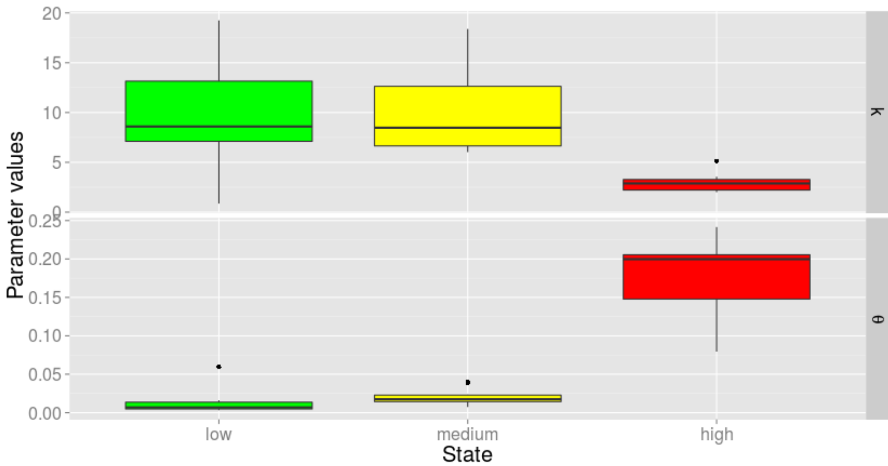


Figure 3.5: Box-plots of the parameters estimates for the state-dependent distributions, note the difference in the scales [Paper C].

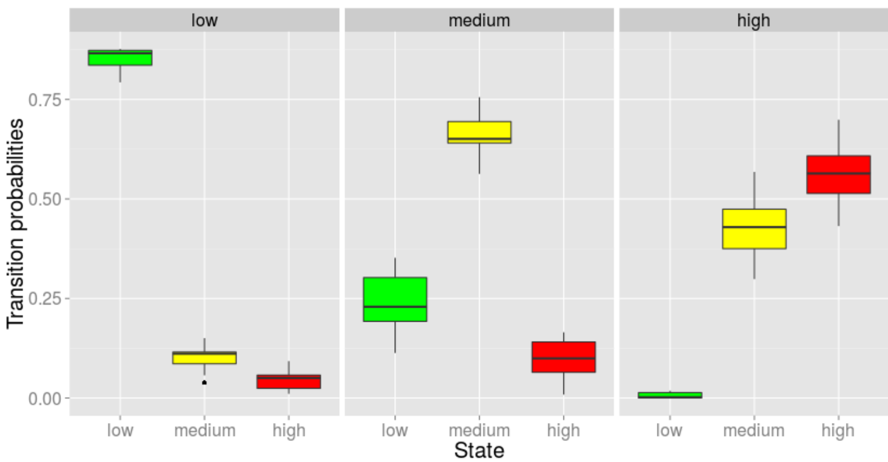


Figure 3.6: Box-plots of the transition probabilities for each state [Paper C].

3.2.4 Inhomogeneous HMM

For the homogeneous models when forecasting more than a couple of hours ahead the probability quickly converges to the stationary distribution. By implementing time dependence in the transition probabilities a great increase in forecasting performance is seen, in Figure 3.7 probabilistic forecast for each hour is shown and it clearly follows the daily pattern observed in Figure 3.4.

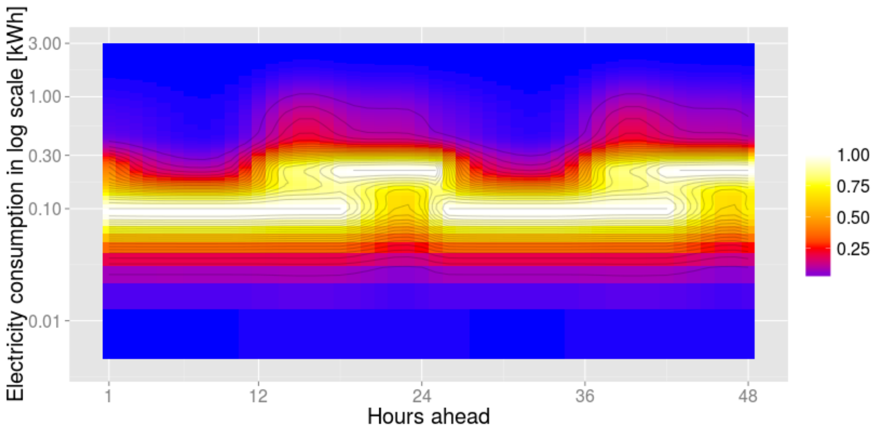


Figure 3.7: Contour plot of forecasting distributions 48 hours ahead of the data used to fit the model. The scale is relative to the largest probability in each horizon, Apartment 18 [Paper C].

Similar observations for common parameters was observed for the inhomogeneous model parameters suggesting population models are possible for these models as well.

A temperature dependent inhomogeneous model was fitted on apartment 2 which also showed an increase in forecast performance but issues with variation of the temperature input needs to be solved.

3.3 Application of Continuous-Time Hidden Markov Models

In Paper D Continuous-Time Hidden Markov Model is applied on one apartment. The data is log transformed to avoid constrained optimization.

3.3.1 Estimated distribution parameters and transition intensities

With the CT-HMM the reduction in the number of parameters makes it possible to find solutions to the optimization problem with many more states than the DT-HMM. For this apartment a nine-state model was found suitable. The parameter estimates are presented in Tables 3.3 and 3.4.

The values in the diagonal of the transition intensity matrix indicates the intensity of changing from the state, if the value is close to zero it is likely to stay, if far from zero it will most likely change. For the off diagonal it is opposite far from zero a change to this state is more likely.

Table 3.3: The parameter estimates of the state distributions, where μ is the mean and σ the standard deviation of the Gaussian distributions. Note that the estimates are based on the log-transformed data.

State	$\hat{\mu}$	$\hat{\sigma}$
1	6.82	0.48
2	5.66	0.26
3	5.13	0.10
4	4.95	0.08
5	5.05	0.06
6	5.26	0.09
7	5.52	0.10
8	5.85	0.12
9	6.12	0.15

Table 3.4: The transition intensity matrix $\boldsymbol{Q}$ [Paper D].

	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
State 1	-0.955	0.208	0.000	0.000	0.000	0.000	0.000	0.000	0.747
State 2	0.337	-0.571	0.234	0.000	0.000	0.000	0.000	0.000	0.000
State 3	0.000	0.620	-0.774	0.154	0.000	0.000	0.000	0.000	0.000
State 4	0.000	0.000	0.366	-0.422	0.056	0.000	0.000	0.000	0.000
State 5	0.000	0.000	0.000	0.443	-0.473	0.031	0.000	0.000	0.000
State 6	0.000	0.000	0.000	0.000	2.569	-3.030	0.461	0.000	0.000
State 7	0.000	0.000	0.000	0.000	0.000	1.963	-2.171	0.208	0.000
State 8	0.000	0.000	0.000	0.000	0.000	0.000	1.001	-1.425	0.424
State 9	0.157	0.000	0.000	0.000	0.000	0.000	0.000	0.618	-0.775

From the transition intensities the transition probabilities for a given time-step can be calculated, in Table 3.5 it is shown for one hour time-step. From both the intensity and the transition probabilities there is an indication that there is a more likely direction in the state change, e.g., going from State 1 to State 4 is more likely through States 9-5 than 2-3.

The sojourn times in Figure 3.8 is also calculated from the transition intensities. These give indications of the time spend in one state. For states with sojourn

Table 3.5: The transition probability matrix $\boldsymbol{P}$, for 1 hour time step. The most likely state transitions are underlined, for each state these will approximately sum up to 91-96% of the transition probability [Paper D].

	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
State 1	$\underline{0.426}$	$\underline{0.103}$	0.012	0.001	0.005	0.008	0.022	$\underline{0.086}$	$\underline{0.338}$
State 2	$\underline{0.167}$	$\underline{0.622}$	$\underline{0.125}$	0.010	0.001	0.001	0.002	0.011	0.061
State 3	0.050	$\underline{0.332}$	$\underline{0.513}$	$\underline{0.088}$	0.003	0.000	0.000	0.002	0.012
State 4	0.007	$\underline{0.065}$	$\underline{0.210}$	$\underline{0.681}$	0.036	0.000	0.000	0.000	0.001
State 5	0.001	0.010	0.048	$\underline{0.289}$	$\underline{0.644}$	0.008	0.001	0.000	0.000
State 6	0.000	0.004	0.025	$\underline{0.191}$	$\underline{0.641}$	$\underline{0.087}$	0.044	0.006	0.001
State 7	0.001	0.001	0.010	$\underline{0.096}$	$\underline{0.477}$	$\underline{0.186}$	$\underline{0.175}$	0.043	0.012
State 8	0.012	0.001	0.002	0.023	$\underline{0.166}$	$\underline{0.129}$	$\underline{0.205}$	$\underline{0.304}$	$\underline{0.157}$
State 9	$\underline{0.071}$	0.008	0.001	0.003	0.030	0.037	$\underline{0.083}$	$\underline{0.229}$	$\underline{0.538}$

time less than one hour is rarely seen more than once in a row, this indicates transition between activities. States with sojourn times around 1 and up to 2 are seen once or twice in the same state before changing, which indicates more interaction with appliances. States with sojourn time higher than 2 are often seen multiple times in a row before changing, this indicates stable consumption with low appliance interaction.

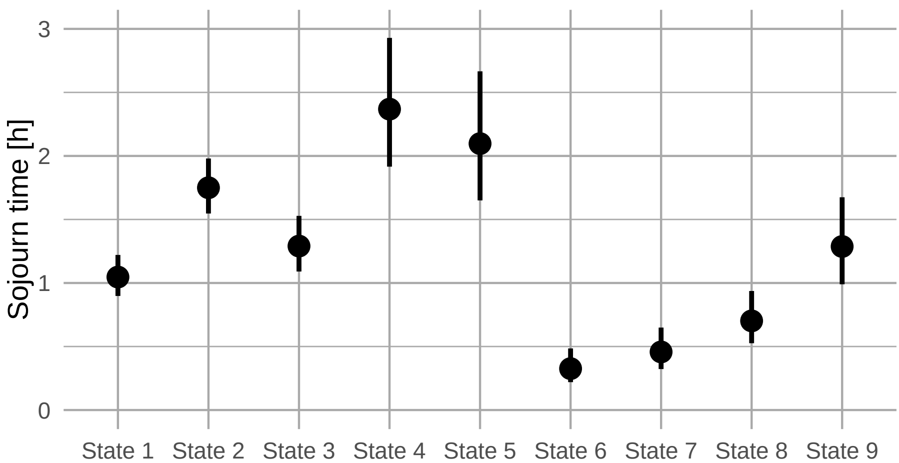


Figure 3.8: The estimated sojourn times of the CT-Markov chain, with 95% confidence band [Paper D].

3.3.2 Global decoding

As for the DT-HMM the most likely sequence of states is found by global decoding. Together with mean values of the underlying distribution the daily variation in the proportion of the states gives indications of possible activity. Compared to the cycle observed in the transition probabilities and intensities we see a clear daily pattern in the cycle through the states, starting from State 5 going through State 4, 3, 2, 1, 9, 8, 7 and 6 suggesting that the states are time dependent and an inhomogeneous CT-HMM could be applied.

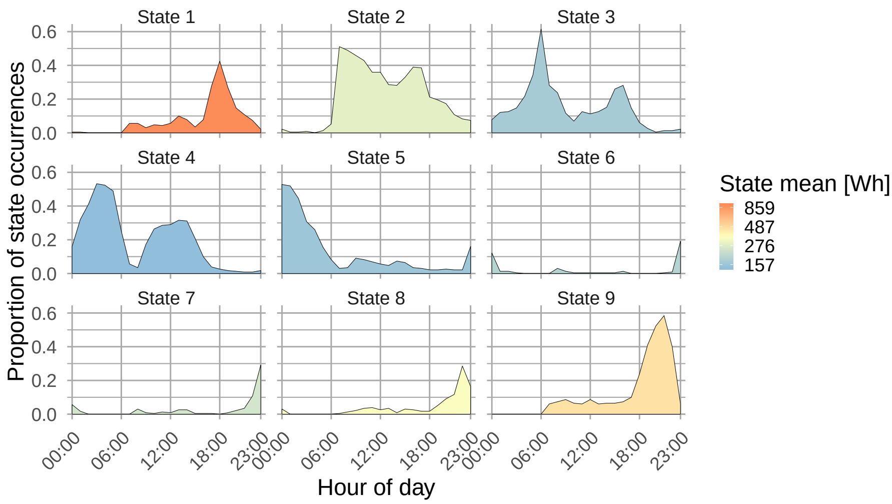


Figure 3.9: The proportion of state occurrences given time of day, coloured by the estimated mean value of the underlying state distribution [Paper D].

3.3.3 State descriptions

By investigating the underlying appliance loads for each state (an example for State 2 is shown in Figure 3.10), together with the proportion profiles, sojourn times, transition intensities, transition probabilities and the underlying state distributions, the following state descriptions in accordance with appliance-related use activity were found reasonable (list from Paper D):

- State 5: Minimum activity, high base load or minor night time activity.

- State 4: Minimum activity, low base load.
- State 3: Ramp up transition state, high base load or minor night time activity.
- State 2: Medium load daytime activity (TV, cooking, medium lighting load)
- State 1: Multiple and/or high load activity (dishwasher, cooking, vacuum etc.).
- State 9: Medium/high load evening activity, (TV, cooking, high lighting load).
- State 8, 7 and 6: Ramp down transition states.

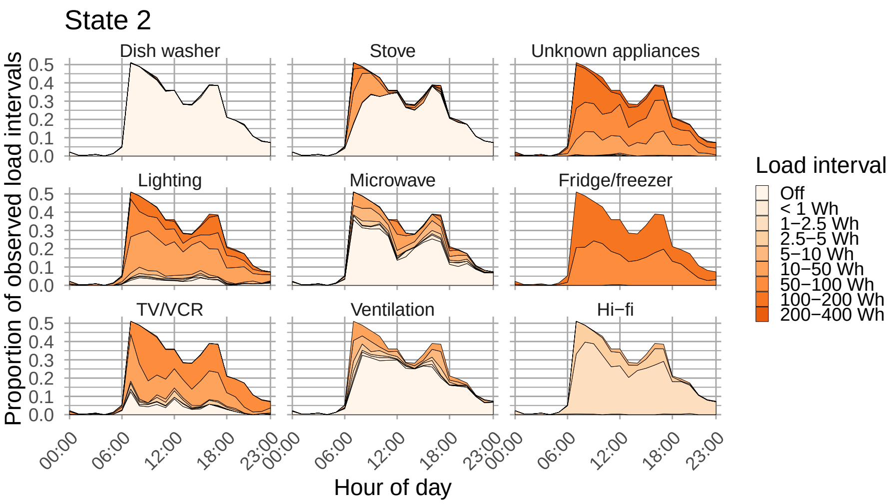


Figure 3.10: The proportion of occurrences for State 2 given time of day, shown with the underlying proportions of observed loads for each appliance time-series [Paper D].

In Figure 3.11 the state descriptions are compared to reported activities from a time-use diary covering four days over a weekend.

A clear correspondence between State 1 and longer cooking activities and turning on the dishwasher is observed.

States 3, 4 and 5 are observed when the residents are sleeping or absent. Work activity can be differentiated into home or elsewhere. If there is an away activity before and after a work period it is elsewhere, otherwise home.

When there are many different activities within the hour, States 2 and 9 are observed. Including shorter cooking activities or activities covering a longer period like watching TV or working at home. It is also observed that State 9 is more likely to occur when both residents are at home doing different things. This might indicate that State 2 is more likely when only one adult resident is at home or a common activity is going on, such as watching TV.

States 6, 7 and 8 are observed when transition from an active to an inactive state is occurring within the hour of observed consumption, which corresponds well to the previous findings.

The state descriptions are corresponding well with the time diary. A ramp down through States 2 and 3 is observed. This is not an unexpected event due to the nature of the CT-HMM, but the ramp down is more likely through States 6, 7 and 8.

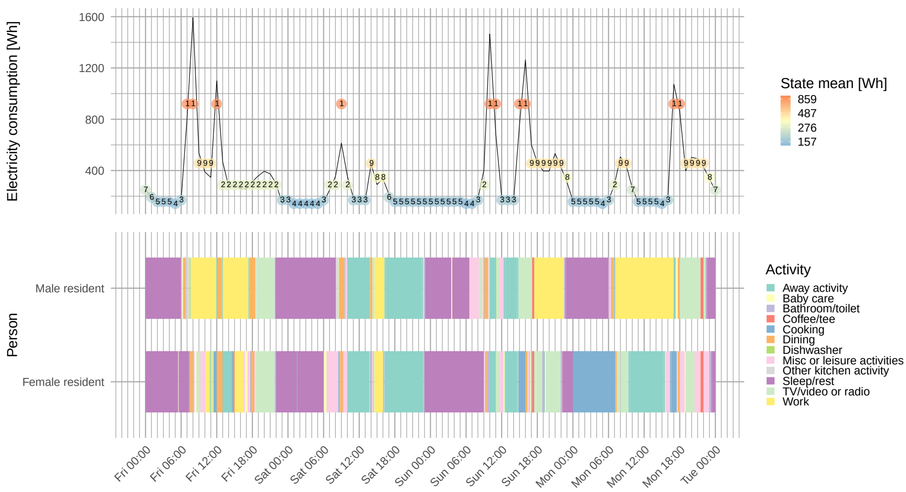


Figure 3.11: Comparison of the most likely sequence of states with time-use diary [Paper D].

3.4 Summary of Hidden Markov Models for disaggregation

The main difference between DT-HMM and CT-HMM is first, the parameter reduction from discrete to continuous time, it is seen that new types of states emerged with increasing the number of states modelled. In the CT-HMM transitions states where discovered and indication of states describing different levels of activity implying different number of people being active in the household. Second, the states can be described based on the parameter estimates, where the DT-HMM is mostly based on the mean of the state distribution the CT-HMM can also rely on sojourn times and the most likely direction to change to.

Chapter 4

Gamification and app-use effect on electricity consumption

In this chapter features in the app Watts is outlined in Section 4.1. The game developed is described and tested in Section 4.2. In Section 4.3 the effect of using the app Watts on electricity consumption is quantified.

4.1 Watts the digital energy assistant

Currently there are three main features in Watts (list from Section E.3):

- (Self-)Monitoring: The main feature in Watts is self-monitoring of the consumption compared to a quarterly forecast.
- Comparison with similar households: A first implementation of comparison between similar household is present in Watts, changes based on the current quarterly estimate of consumption.

- CO_2 monitoring: The current CO_2 load per kWh produced in Denmark is shown and the estimated load for each hour of the day.

For the above features the interactivity is to change the consumption, this is an indirect way of interacting, there is no engaging game design hence the current interaction is hidden. From the current features in Watts we have an effective monitoring tool, but there is not an engaging way of interacting. The reactions to the user's actions are delayed since the smart-meter data is often two days delayed.

4.2 Gamification

Gamification can be defined as: "using game design elements in a non-game context" [36].

The context of electricity consumption we have framed as: "Actions that are meaningful regarding electricity use." [Section E.3]

From the definition of gamification and the notion of procedural rhetoric [37], a goal for gamification of electricity was formulated;

> *We want to increase awareness of the consumers electricity use behaviour (the consumption pattern) triggering beneficial changes in the behaviour, in such a way that the change should be an active choice based on the knowledge gained by playing a game e.g. turning off lights or other appliances or even do energy renovations. This could be done be introducing an engaging way of interacting with electricity use through a game [Section E.4].*

From this goal and other sources of inspiration, a first game concept was developed.

> *Small entities need to be sustained in order to survive and prosper. The entities need to be sustained be the players own electricity consumption, this is done be betting the entities on how the consumption is tomorrow. If the player is correct the entities will prosper otherwise, they will die. By betting on tomorrows consumption, the player can interact with the game by using appliances at a certain time or not, the following day [Section E.4.2].*

4.2.1 The developed game

For the prototype phase of game development, it was not possible to bet into the future and only betting on recent data was possible and the possibility to affect the game using power or not was lost.

The prototype game developed is called PowerMons. The key elements of the game are listed below (list from Section E.4.2);

- PowerMon: An entity that needs electricity to be sustained and multiply. It also represents a charge of the battery.
- The battery: The habitat for the PowerMons to frolic when they aren't assigned to the grid for feeding.
- The Grid: The Grid is where the PowerMons are being placed in order to feed and possible multiply/breed. The Grid is split into 12 columns one for every 2 hours of the day. this makes a square grid of 72 slots where one PowerMon can occupy one slot. each slot represents roughly 5% of the daily power use.
- The Magnet Tool: Since the PowerMons are electricity entities, the idea is that an electromagnet is needed to move them. When the mouse-key is pressed in the Battery all the PowerMons there are attracted to the magnet, multiple PowerMons can be picked up.

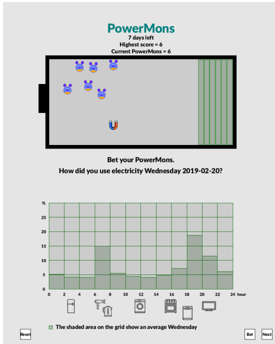
The game is played using the consumers own data from the last 7 days, the goal of the game is to fill the battery in these 7 days be placing the PowerMons in the grid as the consumption was distributed for the given day if placed correctly the PowerMons will multiply. The game play is illustrated in Figure 4.1.

4.2.2 Play testing

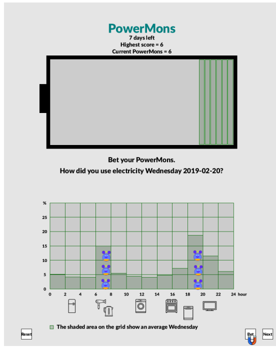
The game was tested with four Watts users, 2 female and 2 male, all employed at SEAS-NVE, working in other departments than the author. The testers had previously tested other features in the app Watts.

In the evaluation of the play test, it is clear that the testers found it difficult to understand how the game worked. The introduction to the game was not good enough and it is suggested to improve the introduction to a tutorial or walk through explaining how the game is played. One of the testers seemed to grasp some of the game play but not all was clear. From the play sessions there is an

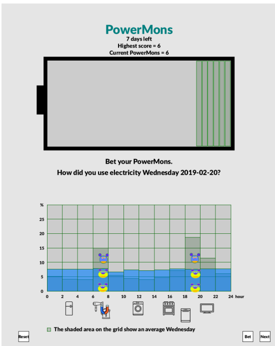
(a) The intro screen briefly describes how the game works.



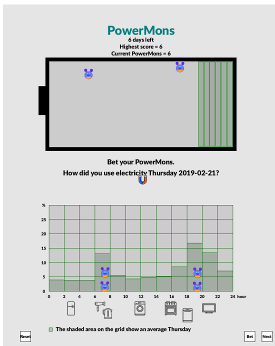
(b) Start of the game. The PowerMons needs to be dragged from the battery down to the grid using the magnet tool.



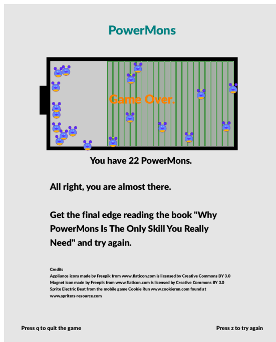
(c) PowerMons are placed in the grid were the player think his consumption was on the day specified above the grid.



(d) The bet button was pressed and the actual consumption was revealed. Two PowerMons were not sustained and will disappear. Since two columns have two PowerMons which are sustained in each of them, two new PowerMons will appear in the battery the next day, one for each column.



(e) The next button was pressed and the game progressed to the next day. Two new PowerMons have appeared in the battery. The sustained PowerMons from the previous day are still placed in the grid and can be moved to new slots.



(f) When the game is over the result is presented to the player with an encouragement to try again if the highest score is not achieved.

Figure 4.1: The game play illustrated [Paper E].

indication of the testers relating their presence in the home to the distribution shown in game.

4.3 App-use effect on electricity consumption

Over time several users have told how they have saved electricity, monitoring their consumption based on the prediction presented in the app.

From late 2017 data on app sessions have been collected, every time the app is opened a record of this is saved. Given this data, it is of interest to investigate if the data can tell a similar story as the users, i.e., if app use have a reducing effect on the electricity consumption and answer the question, is Watts an effective monitoring tool?

To do this linear mixed effect models are utilised.

4.3.1 Data

To investigate a possible effect, first an exploratory data analysis was conducted to find possible explanatory variables. Based on this analysis no clear indication of app use effect was observed but a clear effect on temperature and heating type was observed. The data chosen for the mixed effect model is summarised in 4.1.

Table 4.1: Variable description [Paper F].

Variable	description
Y	The monthly aggregated electricity consumption.
`Appuse`	Count of app sessions within the observed month
`Appusegroup`	Grouping of app use count in intervals.
`Appinstalled`	Was the app installed at the observation
`PrimaryHeatingType`	Primary heating type
`temp.diff`	The temperature difference between fixed threshold (22°C) and ambient temperature.
`InstallationId`	A meter identifier (household)

4.3.2 Effect

Using the data from Table 4.1 as input in a linear mixed effect model, the following model was found best, with no strong evidence against the assumption of normal distributed residuals;

$$\begin{aligned}
\log Y_i = {} & \alpha(\texttt{PrimaryHeatingType}_i) && (4.1)\\
& + \beta(\texttt{temp.diff}_i) && (4.2)\\
& + \gamma(\texttt{PrimaryHeatingType:temp.diff}_i) && (4.3)\\
& + \delta_1(\texttt{Appuse}_i) && (4.4)\\
& + e(\texttt{InstallationId}_i) + \varepsilon_i && (4.5)
\end{aligned}$$

Where $i = 1, \ldots, 97568$,

$$e(\texttt{InstallationId}_i) \sim N(0, \sigma_e^2), \varepsilon_i \sim N(0, \sigma^2) \tag{4.6}$$

all are iid. `PrimaryHeatingType`, `temp.diff` and `Appuse` are fixed effects, and `InstallationId` is the random effects.

Since response variable was log-transformed consumption, we can extract the multiplicative effect of the app use, by using the exponential function on the parameter estimates.

$$AppUseEffect = \exp(\delta_1 \cdot N), \tag{4.7}$$

where δ_1 is the slope parameter for `Appuse` and N is the number of app sessions per month. The multiplicative effects for both `Appuse` and `Appusegroup` are shown in Figure 4.2.

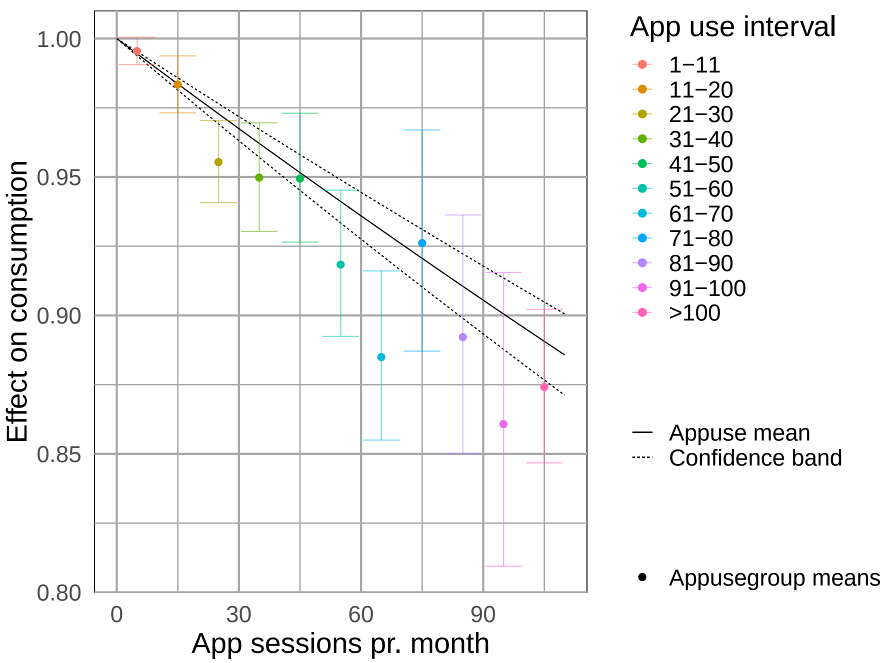


Figure 4.2: The multiplicative effect of app use in single family houses. Multiplicative effect shown with 95% confidence intervals [Paper F]

From Figure 4.2, a significant decreasing effect is seen for increasing app use. When looking at confidence bands for `Appusegroup` it is clear that there is high uncertainty for high number of app use, this is due to the decrease in observations as the N increases, but is also seen that the confidence bands are all below 1.

Chapter 5

Conclusions

5.1 Contributions

In the course of the research described in this thesis, the modelling scheme for long-term prediction of electricity consumption, using adaptive recursive least squares was applied in the app Watts (currently for 45.000 smart-meters). Methods for coarse grained disaggregation has been developed and the effect of using the app Watts has been quantified.

A model procedure for long-term baseline prediction, and results from applying it on daily values of 22,000 households was presented, and model deficiencies where highlighted. Extensions to the modelling framework was implemented based on the deficiencies and with these extensions a potential for great improvement was seen, but this still requires further research. Furthermore, the model procedure shows potential for disaggregating heat load of daily consumption values.

Using DT-HMM on hourly electricity consumption on 14 apartments 3-4 states where found describing the presence of the occupants in the building. Similar daily patterns were observed between apartments and some parameter estimates showed indication that population models are possible. DT-HMM also showed potential for probabilistic short-term forecasting. Using CT-HMM we have successfully applied an approach for coarse grained disaggregation of electricity consumption on a single household. The obtained states of the CT-HMM was described in accordance to appliance related activities using global decoding. The state descriptions corresponded well to a time-use diary. There are indications that the model parameters can describe the states such that the model can be employed as an unsupervised disaggregation approach. The DT-HMM and CT-HMM are yet to be implemented in such a way that the disaggregated feedback can be presented to the users.

A prototype game was developed and tested on four Watts users, using consumers own consumption as part of the game play. Although it was difficult for the players to understand the game in the current state, it gave rise to ideas on how to improve the game and engage consumers through playing.

The effect of the app Watts, on electricity consumption, was found significant when users interact with the app showing that there is a decreasing effect for active users, thus serving as an effective monitoring tool.

5.2 Perspectives and Opportunities for Further Research

For the RLS modelling scheme numerous ideas on how to improve the framework has risen through the real-life application in Watts. By developing metrics describing the volatility and historic changes in consumption and classifying consumption pattern, methods for choosing model types or tuning hyper parameters could be developed such that automation of model type choice and the reasoning for this could be presented to the consumers. Currently a simpler model type choice logic is implemented. By monitoring the evolution of the parameter estimates of the RLS, information on reasons for change might surface this could together with daily load disaggregation be valuable feedback. Investigation of new model structure input could improve the predictive performance and lead to new possible disaggregation.

To use Hidden Markov Models for large-scale disaggregation several things must be in place. First, it must be verified that general state description exists for the CT-HMM. optimization routines must be streamlined and computation time reduced, maybe with population models. Further investigation on the short-term forecasting performance could also be interesting for Home Energy Management Systems. Presenting the findings from the disaggregation to the consumer should be thought of carefully since it is complex information.

It could be interesting to develop a game where the game control partly was electricity consumption or more general energy consumption, if data on heating and water use is available.

Since we only had information on number of app sessions, we cannot answer what in the app aids the decreasing effect. What in the app are leading to the lower effect of active users? Currently data on which pages in the app are used and interacted with are being recorded. With these data we might be able to describe and possible pin point which information that give the lowering effect of app use.

In this thesis we have approached the task of providing feedback/advice to residential electricity consumer through the app Watts, a positive effect was seen for active users. Continuing research on extracting feedback from the data, should consider that it is not enough to fit a model on the data and present the results to the users. The users need the feedback to be easy to understand and interpret in relation to the users own situation.

Bibliography

[1] *What is DataHub?* https://en.energinet.dk:443/Electricity/DataHub. Accessed:2018-02-13. 2018.

[2] *bekendtgørelse om fjernaflæste elmålere og måling af elektricitet i slutforbruget.* https://www.retsinformation.dk/Forms/R0710.aspx?id=160434. Accessed: 2019-04-16. 2013.

[3] *Smart Metering deployment in the European Union.* https://ses.jrc.ec.europa.eu/smart-metering-deployment-european-union. Accessed: 2019-04-17. 2019.

[4] *Data hub.* https://www.svk.se/en/stakeholder-portal/Electricity-market/data-hub/. Accessed: 2018-02-13. 2018.

[5] *2012/148/EU: Commission Recommendation of 9 March 2012 on preparations for the roll-out of smart metering systems.* https://eur-lex.europa.eu/eli/reco/2012/148/oj. Accessed: 2019-04-17. 2012.

[6] R. K. Andersen. "The influence of occupants' behaviour on energy consumption investigated in 290 identical dwellings and in 35 apartments". In: *10th International conference on healthy buildings.* 2012.

[7] C. Fischer. "Feedback on household electricity consumption: a tool for saving energy?" In: *Energy efficiency* 1.1 (2008), pp. 79–104.

[8] P. W. Schultz et al. "Using in-home displays to provide smart meter feedback about household electricity consumption: A randomized control trial comparing kilowatts, cost, and social norms". In: *Energy* 90 (2015), pp. 351–358.

[9] K. Amasyali and N. M. El-Gohary. “A review of data-driven building energy consumption prediction studies”. In: *Renewable and Sustainable Energy Reviews* 81 (2018), pp. 1192–1205.

[10] R. K. Jain et al. “Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy”. In: *Applied Energy* 123 (2014), pp. 168–178.

[11] F. Lai, F. Magoules, and F. Lherminier. “Vapnik’s learning theory applied to energy consumption forecasts in residential buildings”. In: *International Journal of Computer Mathematics* 85.10 (2008), pp. 1563–1588.

[12] Y. Iwafune et al. “Short-term forecasting of residential building load for distributed energy management”. In: *2014 IEEE International Energy Conference (ENERGYCON)*. IEEE. 2014, pp. 1197–1204.

[13] S. Singh and A. Yassine. “Big data mining of energy time series for behavioral analytics and energy consumption forecasting”. In: *Energies* 11.2 (2018), p. 452.

[14] N. Fumo and M. R. Biswas. “Regression analysis for prediction of residential energy consumption”. In: *Renewable and sustainable energy reviews* 47 (2015), pp. 332–343.

[15] K. C. Armel et al. “Is disaggregation the holy grail of energy efficiency? The case of electricity”. In: *Energy Policy* 52 (2013), pp. 213–234.

[16] B. Najafi, S. Moaveninejad, and F. Rinaldi. “Data analytics for energy disaggregation: methods and applications”. In: *Big Data Application in Power Systems*. Elsevier, 2018, pp. 377–408.

[17] J. Kelly and W. Knottenbelt. “Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature”. In: *arXiv preprint arXiv:1605.00962* (2016).

[18] G. W. Hart. *Nonintrusive Appliance Load Data Acquisition Method: Progress Report*. MIT Energy Laboratory, 1984.

[19] G. W. Hart. “Nonintrusive appliance load monitoring”. In: *Proceedings of the IEEE* 80.12 (1992), pp. 1870–1891.

[20] J. Z. Kolter, S. Batra, and A. Y. Ng. “Energy disaggregation via discriminative sparse coding”. In: *Advances in Neural Information Processing Systems*. 2010, pp. 1153–1161.

[21] B. Zhao, L. Stankovic, and V. Stankovic. “Electricity usage profile disaggregation of hourly smart meter data”. In: *4th International Workshop on Non-Intrusive Load Monitoring*. 2018.

[22] B. J. Birt et al. “Disaggregating categories of electrical energy end-use from whole-house hourly data”. In: *Energy and Buildings* 50 (2012), pp. 93–102.

[23] A. Kipping and E. Trømborg. “Modeling and disaggregating hourly electricity consumption in Norwegian dwellings based on smart meter data”. In: *Energy and Buildings* 118 (2016), pp. 350–369.

[24] C. Wang, D. Yan, and Y. Jiang. “A novel approach for building occupancy simulation”. In: *Building simulation*. Vol. 4. 2. Springer. 2011, pp. 149–167.

[25] J. Page et al. “A generalised stochastic model for the simulation of occupant presence”. In: *Energy and buildings* 40.2 (2008), pp. 83–98.

[26] P. D. Andersen et al. “Dynamic modeling of presence of occupants using inhomogeneous Markov chains”. In: *Energy and Buildings* 69 (2014), pp. 213–223.

[27] F. Haldi and D. Robinson. “Interactions with window openings by office occupants”. In: *Building and Environment* 44.12 (2009), pp. 2378–2395.

[28] H. B. Gunay et al. “On adaptive occupant-learning window blind and lighting controls”. In: *Building Research & Information* 42.6 (2014), pp. 739–756.

[29] J. Tanimoto and A. Hagishima. “State transition probability for the Markov Model dealing with on/off cooling schedule in dwellings”. In: *Energy and Buildings* 37.3 (2005), pp. 181–187.

[30] U. Wilke et al. “A bottom-up stochastic model to predict building occupants’ time-dependent activities”. In: *Building and Environment* 60 (2013), pp. 254–264.

[31] D. Johnson et al. “Gamification and serious games within the domain of domestic energy consumption: A systematic review”. In: *Renewable and Sustainable Energy Reviews* 73 (2017), pp. 249–264.

[32] *Watts*. `https://watts.seas-nve.dk/`. Accessed: 2019-05-28.

[33] *SEAS-NVE Om koncernen*. `https://www.seas-nve.dk/koncernen/om/koncernen`. Accessed: 2019-05-28.

[34] H. Madsen. *Time series analysis*. CRC Press, 2007.

[35] S. L. Shah and W. R. Cluett. “Recursive least squares based estimation schemes for self-tuning control”. In: *The Canadian Journal of Chemical Engineering* 69.1 (1991), pp. 89–96.

[36] S. Deterding et al. “From Game Design Elements to Gamefulness: Defining "Gamification"”. In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. MindTrek ’11. New York, NY, USA: ACM, 2011, pp. 9–15. ISBN: 978-1-4503-0816-8. DOI: `10.1145/2181037.2181040`. URL: `http://doi.acm.org/10.1145/2181037.2181040` (visited on 04/18/2017).

[37] I. Bogost. *Persuasive Games: The Expressive Power of Videogames*. en. MIT Press, June 2007. ISBN: 978-0-262-02614-7.

Part II

Publications

Paper A

Models for long-term baseline prediction of daily electricity consumption in individual households

Authors:
Jon Liisberg, Jan Møller, Peder Bacher

Submitted to:
Applied Energy.

Models for long-term baseline prediction of daily electricity consumption in individual households

Jon Liisberg[1 2], Jan Møller[1], Peder Bacher[1]

Abstract

With the increasing use of smart meters, a huge potential has arisen to increase awareness among residential electricity consumers of how they are using electricity and to help them decrease their consumption. To do this, statistical models which can handle the huge diversity in electricity consumption observed in residences can be very useful. The information drawn from the application of such models can be presented to the consumers in an intuitive way and they can easily monitor their consumption, for example through notifications of changes in consumption behaviour. A model can also separate the effect of weather and behaviour.

The present article proposes a modelling procedure composed of a model-selection step and a fitting step, where adaptive models fitted using a Recursive Least squares (RLS) scheme are applied. The procedure is applied on two years of daily electricity consumption data from 22,000 households using reanalysis weather data model input. With the results obtained, the article investigates how well the models perform on predicting daily consumption a quarter of a year ahead in comparison to a benchmark model based on consumption from the same quarter in the previous year. The article then highlights model deficiencies and identifies possible improvements.

A.1 Introduction

With the implementation of the supplier-centric DataHub in Denmark [1], service providers can be allowed by a consumer to access the consumer's smart metering data of electricity consumption. The data available is the total electricity purchase for each meter in hourly resolution. This has a huge potential

[1]DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark
[2]SEAS-NVE, Digital Innovation, Hovedgaden 36, DK-4520 Svinninge, Denmark

to increase awareness among residential electricity consumers on their consumption. In [2], numerous studies were reviewed suggesting that feedback to consumers on their electricity consumption leads to lowering of their consumption. Although these studies use many different approaches, the paper suggests some likely features for successful feedback. A subset of these features is that the feedback is:

- Based on actual consumption.
- Given frequently.
- Given over a longer period.

A more recent study has also provided evidence that feedback can lead to reduction in consumption [3]. This study investigated feedback through in-home displays.

Motivated by this, the present article builds on an idea to use the consumption data as input to a self-monitoring tool, such that consumers can act proactively on the development of their consumption. This is done by predicting consumption over a period and presenting the deviation between the actual consumption and the prediction. If the consumption is below the prediction, consumers get a positive message, and if it is significantly above, they get a warning that their behaviour has changed and is leading to higher consumption.

In order to clarify the use of "predict", the term is defined in the following. We seek to inform the consumer about structural changes that are not related to the weather, but related to some other systematic change in the use of electricity, i.e. a change in behaviour. Furthermore, we want to update this information such that, at any time during the period, the consumer can compare current behaviour with behaviour at the beginning of the period. The approach is to train a model using the observed consumption available up until the beginning of the period and then, during the period, to use the latest updated weather data, so-called reanalysis, as model input to *predict* the baseline consumption. Hence, this prediction indicates what the consumption would have been if the behaviour estimated at beginning of the period had continued during period. The term *estimate* is reserved for estimation of (parameters in) a model. The term *forecast* is used for causal predictions, hence when only data available at the time of forecasting is used - in the present case this would imply that reanalysis weather data could not be used.

Previously, metering data has been used for *long-term forecasts of aggregated consumption* for demand side management [4, 5]. In these studies, segmentation of consumer patterns is used for aggregation and aggregated data is used for forecasting. *Short-term forecast on individual household level* of electricity consumption is also a well-studied area, e.g., [6, 7, 8]. These studies usually have hourly or higher resolution focus on forecasting from hours to days ahead.

In the present paper, we focus on long-term prediction (up to three months ahead) for individual household electricity consumption. Therefore, we looked to find literature presenting studies on mid-term to long-term prediction or forecasting at individual household level. We only found a few studies on this topic. The review paper [9] on data-driven models for forecasting building energy consumption confirms this and concludes that research on long-term forecasting requires more attention. It is indisputable that long-term forecasting poses quite some challenges, as the uncertainty of forecasts is quite high due to the uncertainty of weather forecasts as well as behavioural changes. Of the 63 forecast studies reviewed in [9], three studies were on residential electricity consumption at daily level. In [10], support vector regression for forecasting energy consumption in multifamily residences, both temporal (daily, hourly and 10 minutes) and spatial (whole building, floor and unit) granularity were studied. In [11], support vector machines were used on daily values for forecasting one month ahead: focus was to investigate contribution from explanatory variables. In [12], multiple linear regression was studied for short-term forecasting in individual residences as input to a home energy management system. In these three papers, models for daily forecasting were studied, but not the performance of forecasts on long horizons. Beside these three papers, we found two more studies. In [13], a bottom-up approach using clustering and data-mining of high-resolution data at appliance level was proposed. Based on the patterns found, a Bayesian network was used to forecast hourly, daily, monthly and seasonal resolution. In [14], three different regression methods for hourly and daily consumption were studied. Unfortunately, the paper only evaluates the models in-sample, and no predictive performance was evaluated.

In the present paper, we propose to use time-adaptive models, where parameters and structure are updated as new data arrives. This is because electricity consumption for individual households is very likely to change over time, both due to the general overall increasing trends in electricity consumption and more individual life-changing events e.g. having children, renovations, appliance upgrades, etc. It is therefore unrealistic that a model with constant parameters or even constant structure will be able to fit the consumption data over longer periods of time.

The choice of using reanalysis weather data, hence data which is updated during the prediction period, is the following:

- Forecast weather data could be used, however on horizons longer than approximately 14 days the weather is unpredictable [15].

- Observed weather data could be used, and although it could be more accurate, it is not easily available at many locations.

Reanalysis data is a good choice for the present study, as it gives the best easily available weather information for baseline prediction. In real applications, there is some delay before the reanalysis data is available, and thus some loss in accuracy compared to the presented results must be considered for online applications.

The choice of using daily values is a model-complexity trade-off between, on the one hand, using higher resolution, which would result in unnecessary complexities to be modelled, and on the other hand using, e.g. weekly values, which would not allow frequent enough updates of information for online applications.

In order to evaluate the model's predictive performance, it is compared to a benchmark. The benchmark used is the average daily consumption during the same period one year before. This benchmark is selected because it is the best naive predictor, since it is adapted to each household and easy to implement. The suggested model procedure has two important advantages over this benchmark:

- The model procedure takes the weather conditions during the period into account, whereas the benchmark is based on the weather one year ago.
- Using the model, a comparison with a more recent behaviour can be obtained - behaviour at the beginning of the prediction period, whereas the benchmark will only allow a comparison with behaviour one year ago.

However, it must also be considered that the model procedure has a higher complexity, since a model needs to be selected and estimated individually for each household. Further, the benchmark is more robust in some aspects, e.g. the model cannot handle all patterns and can thus sometimes generate both negative and much too high predictions.

The aim of the present study is to present a model procedure and show that it is potentially useful by evaluating its predictive performance compared to a benchmark. This comparison is used to reveal model deficiencies and suggest possibilities for improving the procedure.

The focus is to investigate a large-scale application of the proposed procedure. The model procedure is based on RLS with forgetting and has been applied for baseline prediction of daily electricity consumption for 22,000 households in each quarter of a year.

The outcome is a model procedure for long-term baseline prediction of daily electricity consumption three months ahead for individual households and there are proposals for further model development.

The outline of this paper is as follows. First in Section A.2 the data used in this study is presented and an exploratory analysis is conducted to choose the explanatory variables used in the modelling procedure. In Section A.3, the suggested procedure is described, and evaluation methods are presented. Then the results of the large-scale application of the procedure are presented and evaluated in Section A.4. Finally, in Section A.5, the results are discussed and proposals for further model development are given.

Nomenclature

RLS	Recursive Least Squares.
RCE	Relative Cumulated Error.
RMSE	Root Mean Squared Error.
SSE	Sum of Squared Error.
SSER	Sum of Squared Error Ratio.
$\mathbf{a}, \boldsymbol{\theta}$	vectors.
$\mathbf{a}^T$	row vector.
$\mathbf{A}$	matrix.
k	prediction horizon.
t	a time stamp in discrete time.
n	maximum of t, i.e. $t \in \{1, ..., n\}$.
$\mathbf{T}, T_t$	a vector of temperature and the temperature at time t.
$i, j \in \mathbb{Z}$	integers.
B	a set of integers.
$\mathbf{A}_{*,B}$	the columns of $\mathbf{A}$ which are in the set B.

A.2 Data

The electricity data used was from households in Zealand, Denmark. The weather data was from ECMWF[3] and is climate reanalysis data [16] for the locations of the individual households. By combining these data sets, time series with daily average values spanning two years were obtained for the following variables:

- Electricity consumption. This is the electricity purchased by the household. If a household has production from a PV or a wind turbine, for example, it is recorded as energy sales, but the production itself is not recorded, hence self-consumption of the production cannot be determined.
- Ambient temperature.
- Solar radiation (surface solar radiation downwards and total sky direct solar radiation at surface).
- Wind speed and direction.
- Atmospheric pressure.

Furthermore, each household reported:

- Location of the household.
- Type (house, apartment, holiday house).
- Heating source both a primary and a secondary.
- Number of residents.
- Building area in square-meters.

A.2.1 Exploratory analysis

First households with the following features were removed: holiday houses, time-series with fewer than 700 days of data, households with more than 50% zero observations and households with extreme outliers. After this, 22,000 households remained and an exploratory analysis was carried out on these to visualize

[3]https://www.ecmwf.int/

important properties of the data - especially to learn about how the variables are related, and thus to get an idea of which variables hold predictive information and could be useful as explanatory variables in the model.

Consumption was grouped on primary heating type and the daily averages for each group are plotted in Figure A.1. There is a clear dependency on time of

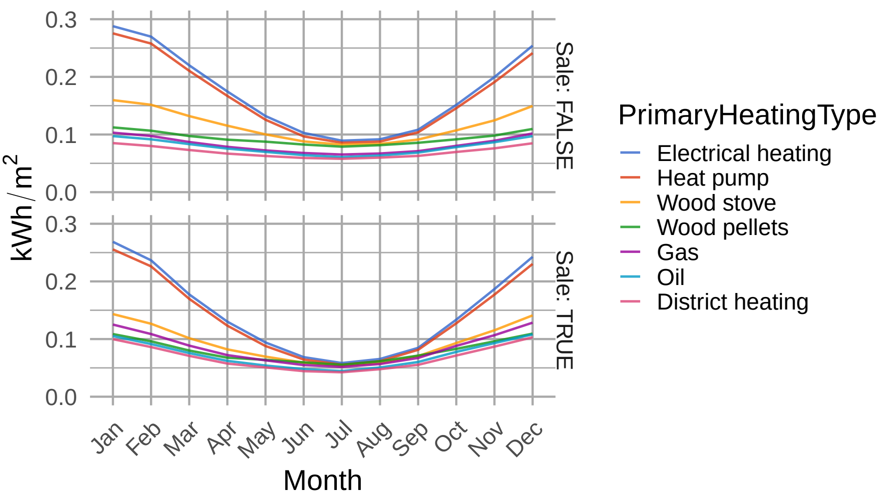


Figure A.1: The daily average electricity consumption per square-meter given primary heating type. The upper panel shows households without electricity production. The lower panel shows households with electricity production.

year - especially for households using electricity for heating. The majority of the households with sale have PV-systems and these are highly affected be the variation of solar radiation over a year. By comparing the upper and lower plots in the figure, a lower summer average for the households with sale can be seen. Furthermore, it can be seen that there is a yearly variation in households without electrical primary heating. This might be due to a secondary heating type or additional lighting and more indoor activity during the winter. These findings indicate that the models should be capable of describing a yearly pattern.

A.2.1.1 Temperature dependence

In Figure A.1, households with electrical primary heating type stand out, hence it is natural to start with the ambient temperature when looking for potential explanatory variables.

For reasons clarified later, the ambient temperature was not used directly as input, but instead the temperature difference between a fixed temperature and the ambient temperature was used. To start out with, a fixed temperature of 22°C was used. Any negative values of this difference, hence values above 22°C were removed. Figure A.2 shows consumption vs. temperature difference. The

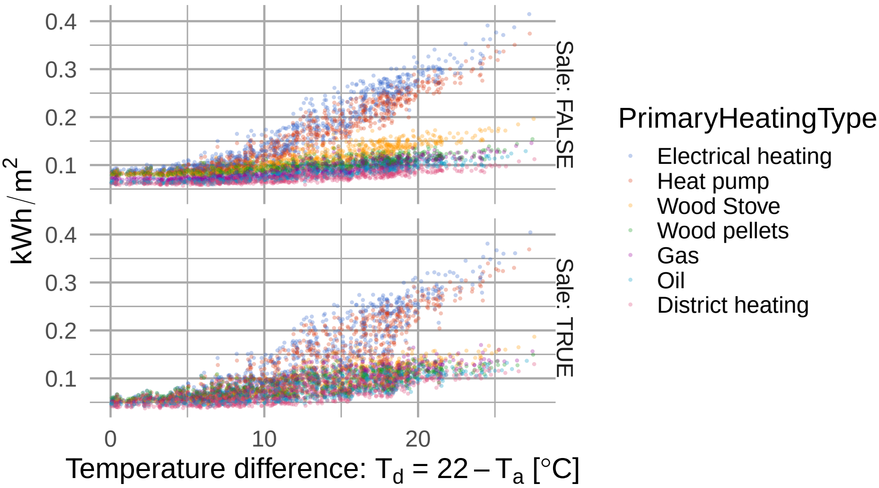


Figure A.2: The daily average electricity consumption per square-meter given primary heating type vs. the temperature difference from 22 °C to ambient temperature. The upper panel shows households without electricity production. The lower panel shows households with electricity production.

strongest relation between the consumption and the temperature difference is clearly for households with electrical heating and heat pumps. A similar, but less steep relation is found for households with other primary heating types. This could be caused by several effects. First, the secondary heating type could be electrical. Second, since the temperature has a yearly pattern, other phenomena with yearly variation could explain this. For the households with sale, there seems to be a higher variation in the electricity consumption compared with the households without sale (again note that electricity consumption is the actual electricity purchase - hence electricity from the households' PV is not recorded in the purchase). This could indicate that other model inputs are needed in these cases. For households with electrical primary heating, there seems to be a change in the slope at around 15°C. This indicates that there might exist a temperature threshold explaining the shift from heating season to off-heating season.

The fixed temperature level used when calculating the difference will later be optimised for each individual household in order to describe the threshold for turning heating on and off. For households with electrical heating, the threshold is observed around 7°C temperature difference in Figure A.2.

A.2.1.2 Solar radiation

Since many households have PV systems to produce electricity, solar radiation data was investigated.

In Figure A.3, the daily average electricity consumption per square-meter is plotted against solar radiation. It is clear that there is a dependence between

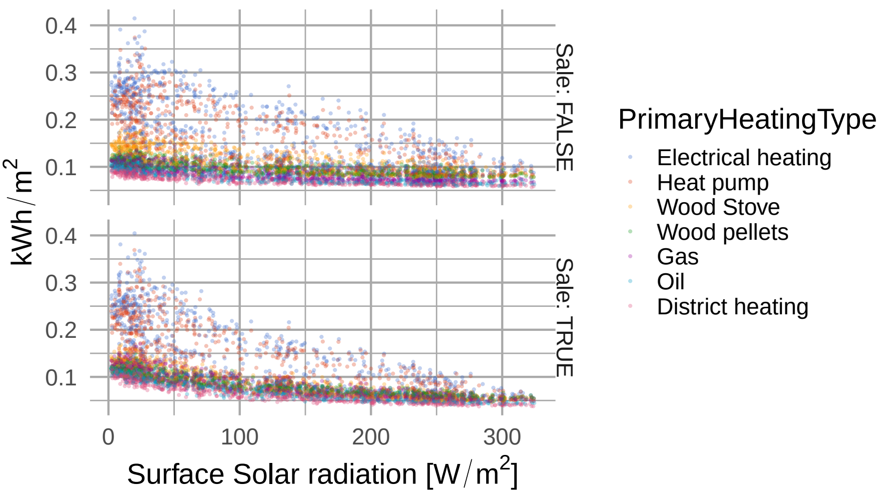


Figure A.3: The daily average electricity consumption per square-meter given primary heating type. The upper plot shows households without electricity production. The lower plot shows households with electricity production.

increasing electricity consumption and decreasing solar radiation in households with electricity sale. As for the temperature, we also observe a similar, but smaller effect for households without sale. For households with electrical heating, there is also an increase in consumption when solar radiation decreases, but with much higher variation, hence this could be due to the correlation between decreasing temperature and solar radiation.

Since we are particularly interested in the precise effect of the solar radiation on the electricity consumption each day, other variables should be considered describing the yearly variation. The number of hours of darkness could work as a viable alternative to radiation. It could also be a variable explaining the yearly variation in households without sale and electrical heating. Another alternative could be to use cloud cover and calculate the clear sky radiation based on location of the household. Figure A.4 shows a similar relation in electricity consumption for increasing number of hours of darkness as was found for decreasing solar radiation. Furthermore, for households without sale, a periodic

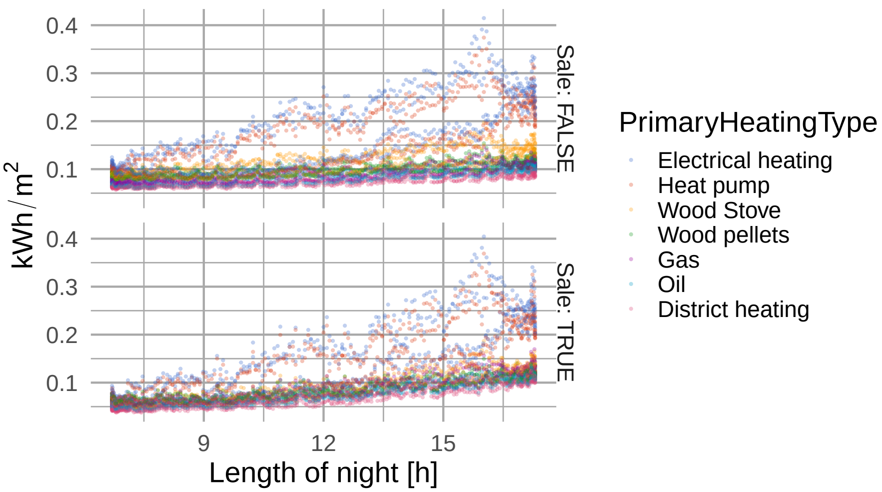


Figure A.4: The daily average electricity consumption per square-meter given primary heating type. The upper plot shows households without electricity production. The lower plot shows households with electricity production.

pattern is observed for non-electrical heating. The period length of this corresponds well to a week, which indicates inputs to describe weekly variation are needed.

At the time when the modelling was carried out, it was decided not to use solar radiation, but only the number of hours of darkness. This was simply because the solar radiation would not be available in the data for the online application planned. Hence, unfortunately, the potential for improvement of the predictions using solar radiation was not investigated further in this study, but should in future studies.

A.2.1.3 Weekly variation

An indication of weekly variation for many households was found in Figure A.4. To further investigate this, the average consumption for each week day is shown in Figure A.5. This shows that weekly variation is present for all heating types.

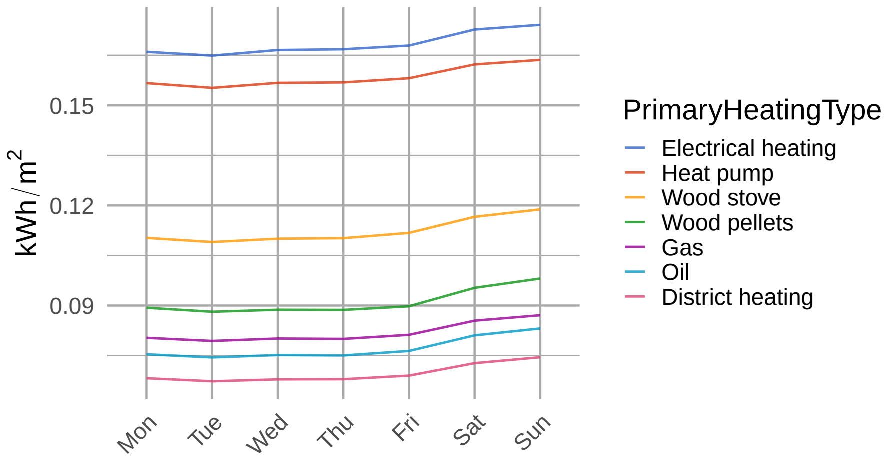


Figure A.5: The daily average electricity consumption per square-meter given primary heating type

Since this is averaged over 22,000 households, naturally for each household the weekly pattern can be very different from the general pattern seen in the plot. Hence an input capable of describing many different patterns in needed. This can be handled by Fourier series as described in Section A.3.2.

A.2.1.4 Selection of explanatory variables

A similar explanatory analysis was carried out for wind speed and wind direction, but neither indicated any significant influence on the aggregated averages, although this might not be true for individual households. Wind can have a significant effect for electrical heating consumers [17] and some even have a household wind turbine which will affect consumption. However, since there are only very few, wind was disregarded as an input.

Since most buildings in Denmark are well insulated and built with bricks and other thermally heavy materials, a low pass filtering of the temperature difference was carried out to describe a dynamic effect and thus investigate whether this will explain the influence of the temperature better than the raw signal. This low pass filtering was carried out as in [18]. Based on the exploratory data analysis, the following model output and input variables were selected to test the model procedure:

- Output variable:
  - Electricity consumption.
- Input variables:
  - Temperature difference between a temperature threshold and the daily average ambient temperature.
  - Low pass filtered temperature difference to model the envelope of the house.
  - Number of hours of darkness given location of the household.
  - Fourier series for describing the weekly variation.

A.3 Methods

This section presents the model procedure. The purpose of the procedure is to find and tune a suitable model for each individual household. In the first stage, a model structure - which explanatory variables to include for a household - is determined using a forward selection approach to test of overall dependence on input variables. In the second stage, optimization of the RLS for each individual horizon is carried out by choosing the harmonics in Fourier series describing the weekly pattern.

A.3.1 Stage 1: Model selection, temperature threshold identification and temperature filtering

A.3.1.1 Model selection using a forward selection procedure

To choose which explanatory variables should enter the model for each household, a linear regression model is fitted using forward selection with the Bayesian Information Criteria (BIC) as the model fit criterion. The BIC is a measure of the relative fit of a statistical model to a given set of data. A lower BIC value indicates a better fit of a model to the given data set. The forward selection works by starting with a null (intercept) model and in each step adding the variable that will improve (decrease) the BIC value the most, until no more improvement is possible. To keep things simple in this stage, the cross-product terms or higher order terms of the explanatory variables in the linear regression forward selection scheme are not considered. If temperature dependence is found during forward selection, then either the temperature difference is chosen or the filtered temperature difference is chosen, not both.

When the best model has been selected for each household, the estimated parameters are checked. They must all, except the intercept parameter, be positive in order to be in line with the physical effects, for example, from a physical point of view, a higher temperature difference should result in a higher consumption. Hence negative correlations to the selected explanatory variables do not make sense and are therefore removed from the model for that household (see Sections A.2.1.1 and A.2.1.2 for descriptions).

A.3.1.2 Low pass filtering of temperature difference

In order to describe the change from heating season to off-heating season, and to obtain the desired parameter for direct interpretation (see Section A.2.1.1 for description), the temperature difference $T_{\mathrm{d},t}$, between an arbitrary fixed threshold $T_{\mathrm{threshold}}$, and the ambient temperature $T_{\mathrm{a},t}$, at the time t is calculated by

$$T_{\mathrm{d},t} = \begin{cases} T_{\mathrm{threshold}} - T_{\mathrm{a},t} & \text{if } T_{\mathrm{a},t} < T_{\mathrm{threshold}} \\ 0 & \text{else} \end{cases}, \tag{A.1}$$

this is done for each individual household such that the optimal individual threshold is found which minimizes the BIC value of a linear regression model, only varying the threshold.

As mentioned previously, buildings in Denmark are thermally heavy and there will be a dynamic effect from changes in the ambient temperature to the resulting changes in heat consumption. In other words, when the ambient temperature changes, the effect on electricity consumption is "delayed" due to the resistance in the envelope (the insulated walls) and the heat capacity of the house (the heat stored in the interior walls and the furniture). Therefore, low-pass filtering is applied to the temperature difference input to be able to describe the dynamic effect with the model. A discrete time, first order RC low-pass filter with a stationary gain of 1 is given by

$$T_{\mathrm{f},i} = a_{\mathrm{Ta}} T_{\mathrm{f},i-1} + (1 - a_{\mathrm{Ta}}) T_{\mathrm{d},i}, \tag{A.2}$$

where $T_{\mathrm{f},i}$ is the filtered temperature difference $T_{\mathrm{d},t}$ and a_{Ta} is a smoothing factor corresponding to the time constant for the part of the building affected by changes in ambient temperature. An appropriate value for hourly time-steps ($a_{\mathrm{Ta},1}$) is obtained from [17], where a similar low-pass filter was applied and the smoothing factor was optimized for individual buildings. Since this smoothing factor is for hourly time steps, we need to calculate it for daily time steps.

The time constant (denoted RC) is the same for all sampling periods and with a known smoothing factor it is given by

$$RC = \Delta \frac{a_{\mathrm{Ta}}}{1 - a_{\mathrm{Ta}}}. \tag{A.3}$$

Rearranging this equation, we obtain an equation for the smoothing factor,

$$a_{\mathrm{Ta}} = \frac{RC}{RC + \Delta}. \tag{A.4}$$

Here, Δ denotes the sampling period. In this setting, the hourly time steps have sampling period $\Delta_1 = 1$ and the daily time steps have sampling period $\Delta_{24} = 24$. By inserting the RC expression (Equation A.3), with the hourly sampling period in Equation (A.4) with daily sampling period we have Equation (A.5). This can be reduced to Equation (A.6) and the daily smoothing factor can be calculated with the hourly by

$$\begin{aligned} a_{\mathrm{Ta},24} &= \frac{\Delta_1 \frac{a_{\mathrm{Ta},1}}{1-a_{\mathrm{Ta},1}}}{\Delta_1 \frac{a_{\mathrm{Ta},1}}{1-a_{\mathrm{Ta},1}} + \Delta_{24}} && \text{(A.5)} \\ &= \frac{\Delta_1}{\Delta_1 + \Delta_{24}/a_{\mathrm{Ta},1} - \Delta_{24}}. && \text{(A.6)} \end{aligned}$$

A.3.1.3 Pseudo code stage 1

To summarise the first stage of the model selection scheme it is presented as pseudo code in Algorithm 3. See the nomenclature for notation.

Algorithm 3 Choosing explanatory variables using forward selection

Input:
$\mathbf{y}$, training data model output of length n.
$\mathbf{X}$, full training data input, all possible variables.
Output: $\mathbf{X}_{\text{RLS}}$, the reduced training data input for RLS.

Set $T_{\text{threshold}} = 22$
$\mathbf{T}_d \vee \mathbf{T}_f$ by Equation (A.1) and (A.2)
$\hat{\boldsymbol{\theta}}, \mathbf{X}_{\text{Reduced}} := \textit{Forward.selection}(\mathbf{y}, \mathbf{X})$
if $\mathbf{T}_d \vee \mathbf{T}_f \in \mathbf{X}_{\text{Reduced}}$ **then**
  $\underset{T_{\text{threshold}}}{\arg\min}(\textit{BIC.LinReg}(\mathbf{y}, \mathbf{X}_{\text{Reduced}}))$ for $T_{\text{threshold}} \in [5, 22]$,
end if

$A := \{1, \ldots, \dim(\hat{\boldsymbol{\theta}})\}$, A is the set of indices of the parameter vector.
$B := \{i \in A | \hat{\theta}_i > 0 \vee i = 1\}$, B is a subset of A with intercept ($i = 1$) and where $\hat{\boldsymbol{\theta}}$ have positive values.
$\mathbf{X}_{\text{RLS}} := \mathbf{X}_{\text{Reduced},*,B}$, The input matrix is reduced to the columns with indices in B.

return $\mathbf{X}_{\text{RLS}}$

A.3.2 Stage 2: Recursive Least Squares fitting scheme

RLS with forgetting is a parametric model where the implementation is based on [19]. To briefly give an idea of the RLS algorithm, it is shown for horizon k

Update step:

$$\mathbf{R}_{t-k} = \lambda \mathbf{R}_{t-k-1} + \mathbf{x}_{t-k}\mathbf{x}_{t-k}^T, \tag{A.7}$$

$$\hat{\boldsymbol{\theta}}_{t-k} = \hat{\boldsymbol{\theta}}_{t-k-1} + \mathbf{R}_{t-k}^{-1}\mathbf{x}_{t-k}(y_{t-k} - \mathbf{x}_{t-k}^T \hat{\boldsymbol{\theta}}_{t-k-1}). \tag{A.8}$$

Prediction step

$$\hat{y}_{t|t-k} = \mathbf{x}_t^T \hat{\boldsymbol{\theta}}_{t-k}. \tag{A.9}$$

Here $\hat{\boldsymbol{\theta}}_t$ is the estimated parameter vector, y_t is the observation, $\hat{y}_{t|t-k}$ is the prediction, $\mathbf{x}_t$ is the vector of observed input,$\mathbf{R}_t$ is the inverse sample covariance matrix up to a constant and λ is the forgetting factor. One model for each horizon k is fitted, hence, for a quarterly prediction, we need approximately 92 model fits for each household depending on quarter and year.

A.3.2.1 Choosing Fourier coefficients using optimization

In the second stage a model is fitted for each horizon, where harmonics and the forgetting factor are chosen based on optimization using root mean squared error ($RMSE_k$) as the objective function. This is to get the best prediction of the mean, when minimizing $RMSE_k$ [19],

$$RMSE_k = \sqrt{\frac{\sum_{t=k}^{n}(y_t - \hat{y}_{t|t-k})^2}{n-k}}, \tag{A.10}$$

where k is the horizon and n is the number of observations of the training period.

Weekly curve In order to capture the weekly pattern observed for many households in the model (mainly the difference between weekends and weekdays), Fourier series are applied as in [20]. The parameter vector $\hat{\boldsymbol{\theta}}_t$ is split into two parts $\hat{\boldsymbol{\theta}}_t^p = (\hat{\boldsymbol{\alpha}}_t, \hat{\boldsymbol{\beta}}_t^p)$, where $\hat{\boldsymbol{\alpha}}_t$ is the vector of explanatory variables parameter coefficients found suited in the forward selection. $\hat{\boldsymbol{\beta}}_t^p$ is the vector of p pairs of Fourier coefficients. For $p = 0$, we define $\hat{\boldsymbol{\theta}}_t^0 = \hat{\boldsymbol{\alpha}}_t$, if no weekly pattern is needed for the household. The Fourier series is given by

$$\sum_{i=1}^{p} \beta_{i_{\sin}} \sin\left(\frac{i2\pi t}{r}\right) + \beta_{i_{\cos}} \cos\left(\frac{i2\pi t}{r}\right), \tag{A.11}$$

where t is the time, r the period and p is the number of sin- cosine pairs.

Optimization routine The model chosen is based on optimization using root mean squared error (Equation (A.10)) as the objective function. For each horizon k, models are fitted choosing the number p of sine and cosine pairs

$$\arg\min_{p}(RMSE_k(\mathbf{y}, \hat{\mathbf{y}}^p)), \text{ for } p \in \{0, ..., 3\}, \tag{A.12}$$

where $\hat{\mathbf{y}}^p$ is a vector of consumption obtained from the prediction step (Equation (A.9)) of the model fitting with the parameter estimates $\hat{\boldsymbol{\theta}}_t^p$. The maximum number of sin and cosine pairs is set to 3, since this should be sufficient to describe a weekly curve with variation between days (6 parameters for 7 days).

Due to optimization issues the forgetting factor λ has been fixed to 0.999, which gives a relative high memory (weighting is halved in approximately 700 days).

The forgetting factor could be a part of the optimization routine to tune the parameter adaptivity to the individual household, but this requires additional work on bounding the inverse covariance matrix $\mathbf{R}$ to avoid computational overflow or underflow [21].

A.3.2.2 Pseudo code stage 2

To summarise the second stage of the model selection scheme, it is presented as pseudo code in Algorithm 4. See the nomenclature for notation.

Algorithm 4 RLS modelfit

Input:
$\mathbf{y}$, training data model output of length n.
$\mathbf{X}_{\mathrm{RLS}}$, training data input.
k, the prediction horizon in number of days.
p, number of sine, cosine pairs of the Fourier series input.
$\mathbf{X}_p$, matrix of p pairs of Fourier series training data.
Output: *ModelList*, list of model fits of length k.
for $i = 1$ **to** k **do**
  for $j = 0$ **to** p **do**
    $\mathbf{X} := [\mathbf{X}_{\mathrm{RLS}}\mathbf{X}_p]$, Fourier series input is combined with the input found in Stage 1.
    $\hat{\boldsymbol{\theta}}_n^j, \hat{\mathbf{y}}^j := RLS(\mathbf{y}, \mathbf{X})$, The RLS is fitted for each added pair of Fourier series input.
  end for
  $q := \arg\min_j(RMSE_i(\mathbf{y}, \hat{\mathbf{y}}^j))$, The fit with smallest RMSE is found.
  ${}_i\hat{\boldsymbol{\theta}}_n := \hat{\boldsymbol{\theta}}_n^q$
end for
return $ModelList := \{{}_1\hat{\boldsymbol{\theta}}_n, \cdots, {}_k\hat{\boldsymbol{\theta}}_n\}$

The *Modellist* is then used with Equation (A.9) and prediction of the input to calculate the prediction of the model output.

A.3.3 Evaluation measures

In order to understand the predictive performance of the model, it must be tested on a large set of households. Two performance measures are used to compare the predictions from the model vs. the benchmark, as well as between the groups of different heating types. As mentioned before, the benchmark is the average consumption from the same period (i.e. same quarter one year ago)

$$\hat{y}_{i_{\text{benchmark}}} = 1/k \sum_{j=1}^{k} y_{n-365+j} \text{ for } i \in \{1, \cdots k\}. \tag{A.13}$$

In the evaluation, we seek mainly to answer the following:

- What is the predictive performance of the model vs. the benchmark? The model should be able to provide a more accurate prediction, since the weather data during the prediction period is used, and more recent data is used compared to the benchmark.
- What can be learned from the evaluation and taken into account to improve the model procedure in future developments?

The first measure is a relative measure of prediction bias. The Relative Cumulated Error (RCE)

$$RCE = \frac{\sum_{i=1}^{k}(\hat{y}_{n+i|n} - y_{n+i})}{\sum_{i=1}^{k} y_{n+i}}, \tag{A.14}$$

where n is the number of observations used for model fitting and k the prediction horizon. Since it is a relative measure, it is possible to compare the performance of the model to the benchmark directly between the households. Hence, it is easy to make groupings based on parameters that have entered the model, or prior knowledge of the household like heating type and compare them in order to identify model deficiencies and possible improvements using the RCE. Patterns are analysed, e.g., if a clustering reveals a bias for a particular cluster, then maybe something can be learned about model deficiencies and possible improvements.

To validate the prediction accuracy on daily values, the Sum of Squared Errors Ratio (SSER) between the prediction SSE and the benchmark SSE

$$SSER = \frac{\sum_{i=1}^{k}(y_{n+i} - \hat{y}_{n+i|n})^2}{\sum_{i=1}^{k}(y_{n+i} - \hat{y}_{i_{\text{benchmark}}})^2}. \tag{A.15}$$

is used. This is also a relative measure and direct comparisons between households can easily be carried out. If the SSER is less than one, then the RLS model is performing better than the benchmark.

A.4 Results

The modelling scheme was applied on almost 22,000 times-series of household electricity consumption. For each household, data was available from all of 2014 and 2015. For each quarter of 2015, a model using data one year before the beginning of the quarter was selected and fitted for each household - thus quarterly predictions (90-92 days ahead) were obtained for four quarters.

The models gave negative predictions for some of the households; the exact number for each quarter is given below. These were removed in the preceding evaluation. Furthermore, households with less than 60 kWh electricity consumption in the period of the prediction or in the same period in the last year were also removed, since this indicates long absence in the periods and comparing with such a period of unusually low consumption is not of interest.

Counts of the different model structures selected in the forward selection procedure are shown in Table A.1. From these it is clear that there is an issue with

Table A.1: Counts of the selected model structures in each of the four predicted quarters. The structure describes which input variables were chosen in the forward selection procedure. A 1 denotes the intercept, LN denotes number of hours of darkness, T_{d} is the temperature difference and T_{f} is the filtered temperature difference. The Fourier input is not shown.

Quarter	1	2	3	4
Input structure				
(1)	1597	2096	1809	1712
$(1, LN)$	5644	5685	5982	5735
Subtotal	(7241)	(7781)	(7791)	(7447)
$(1, T_{\mathrm{f}})$	4218	3343	3139	2991
$(1, T_{\mathrm{f}}, LN)$	7707	7612	6963	7881
$(1, T_{\mathrm{d}})$	677	703	771	756
$(1, T_{\mathrm{d}}, LN)$	1529	1787	2120	2344
Subtotal	(14131)	(13445)	(12993)	(13972)
Total	21372	21226	20784	21419
Reported electrical heating	9009	8905	8642	9013
Removed ($\hat{y}_i < 0$)	436	618	1061	424

negative values ($\hat{y}_i < 0$) for one or more of the horizons in the prediction, particularly for the third quarter. We also observe a much lower number of reported electrical heating than the number of model fits where temperature dependence was found. Again, this can be due to secondary electrical heating or correlations between temperature and behaviour. If this behaviour is systematic, it should be kept in the model, since we are seeking to identify changed behaviour.

A.4.1 Predictive performance

To validate the predictive performance, the RCE in Equation (A.14) is considered. Since the RCE is a measure of bias, then the mean and median of the distribution of RCE over all the households should both be zero (if all accumulated systematic effects are well modelled). However, for the benchmark, which is based on last year's consumption ($\text{RCE}_{\text{benchmark}}$), a non-zero bias should be expected for temperature-dependent households, as a function of the differences in weather conditions between the years: If the quarter was colder than the same quarter last year, then the mean and median RCE should be negative (negative bias because of under-prediction) and vice versa.

The empirical distributions of the model predictions and benchmark RCE are compared using box-plots of the RCE. In Figure A.6, the empirical distributions of the RCE calculated for all the households are shown as box-plots. It is quite

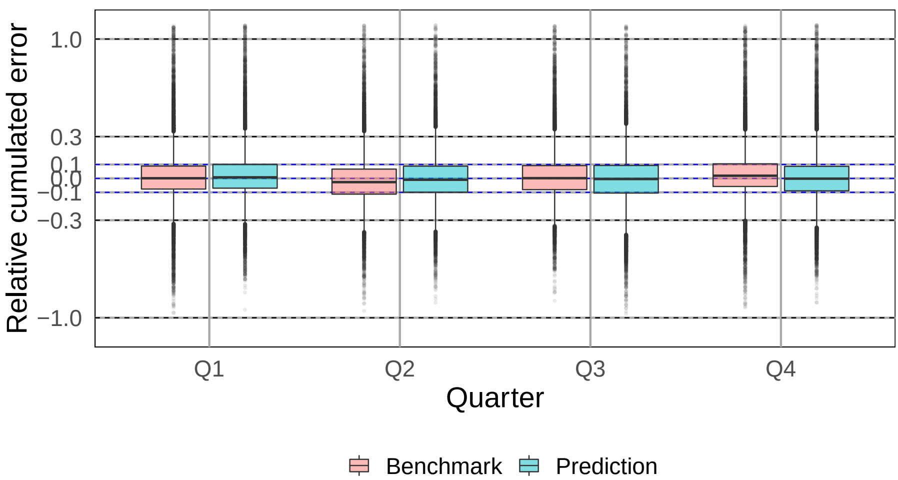


Figure A.6: The empirical distribution of the RCE for all households for each predicted quarter, both for the benchmark and the model predictions.

clearly seen that there is very little difference between the distributions for the benchmark and the model predictions. Also, they seem quite symmetrical and centred around zero, hence no noticeable bias was found.

Going a little more into detail, the same type of box-plots is generated for the groups of selected model structures - the groups listed in Table A.1. They are shown in Figure A.7, again with the benchmark and model next to each other in each quarter. In general, some variation between the distributions of the six

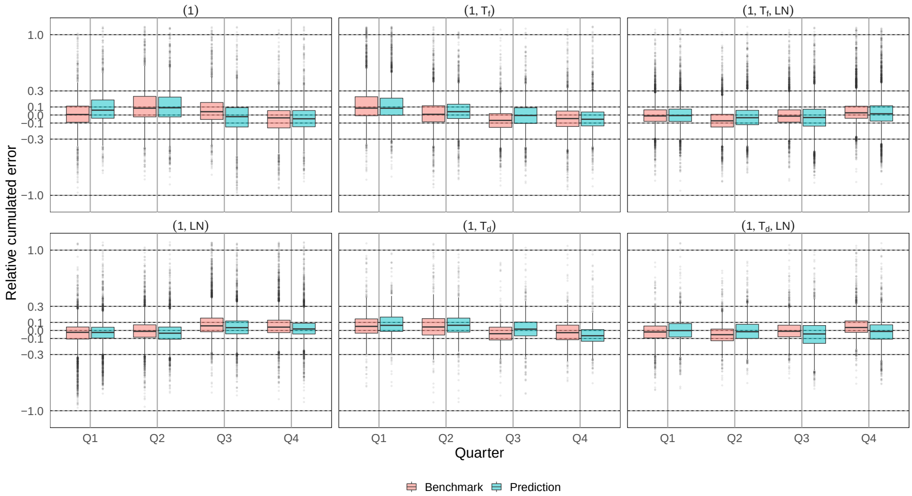


Figure A.7: The empirical distribution of the RCE of the model prediction compared to the RCE of the benchmark. Each plot shows the distribution of the RCE given the selected model structure indicated above the plot.

groups can now be seen, and overall it seems that the benchmark and the model predictions share the same tendencies - meaning that they are under- and over-predicting in the same quarters. Considering the temperature-dependent model structures $(1, T_\mathrm{f})$ and $(1, T_\mathrm{d})$ (the two plots in the middle), the most noticeable difference is seen in Q3, hence in the summer quarter, where in Denmark there should not be a huge difference, since cooling is quite rare. These results lead to the finding that the model predictions are in general not very different, in terms of bias, from the benchmark.

To investigate the impact of temperature, which, as explained earlier, should give rise to some differences in bias between the benchmark and the model predictions, box-plots of the daily temperatures in all of the quarters of the two years are shown in Figure A.8. Thus, for Q1 and Q2, some negative bias could be expected for the benchmark (i.e. $\mathrm{RCE}_\mathrm{benchmark} < 0$. For Q3, the summer period, no noticeable bias could be expected. For Q4 a small positive $RCE_\mathrm{benchmark}$ could be expected.

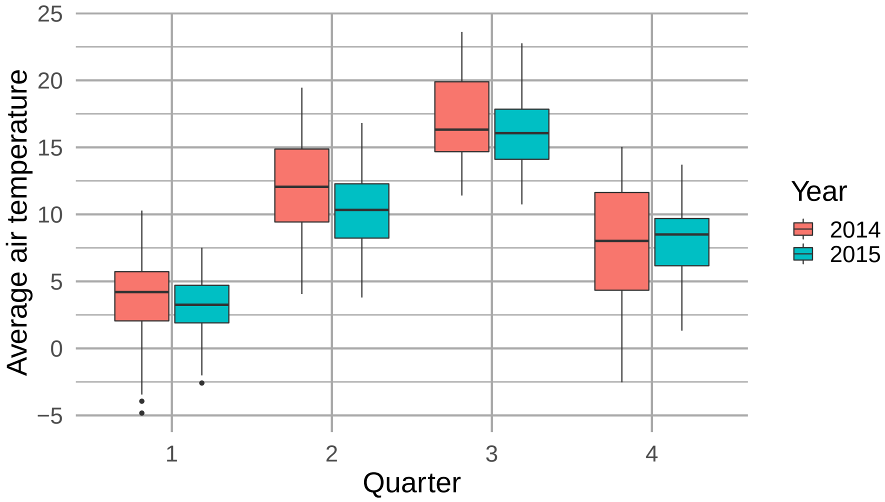


Figure A.8: Comparison of daily average temperatures between 2014 and 2015 for each quarter.

Looking back to the RCE box-plots in Figure A.7 for the temperature-dependent model structures (the four right-most plots) the expected negative bias for the benchmark in Q1 and Q2, and positive bias in Q4 are not at all seen clearly - almost an opposite tendency appears. Most likely there was not a significant difference in temperatures between the two years, however, since only two years are compared it was not found possible to conclude upon this.

A.4.2 Daily predictions

To analyse the prediction performance at a daily level, the SSER in Equation (A.15 is calculated for each household in each quarter and grouped on the selected model structure. Box-plots of empirical distribution of these are shown in Figure A.9. If the SSER is smaller than one, then the model performs better than the benchmark. There are a few cases were the Benchmark SSE is almost zero, which gives rise to a very high SSER value, these are excluded from the plots.

For the intercept model and the model with only number of hours of darkness as input (the two left-most plots), the median is slightly under one, indicating no noticeable differences in prediction accuracy between the benchmark and the model predictions.

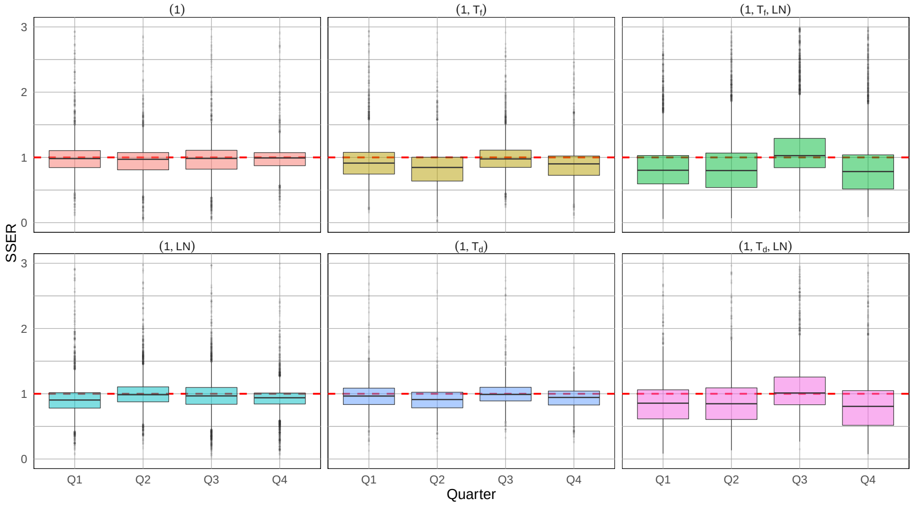


Figure A.9: The empirical distribution of the SSER, given model structure and quarter. Each plot shows the distribution of the SSER given the selected model structure indicated above the plot.

For the temperature-dependent model structures (the two middle plots), the median is below one for all quarters, hence the model predicts in general more accurately than the benchmark of daily levels.

Finally, for the temperature- and number-of-hours-of-darkness-dependent model structures (the right-most plots) a clear general improvement of model over the benchmark is found, except for Q3, which is meaningful, since this is the summer quarter and thus the temperature should not impact the predictions.

A.4.3 Analysis of selected examples

This section presents an analysis focusing on examples posing challenges to the model and what leads to differences between the benchmark and the model predictions. The aim is to give an idea of what leads to either the benchmark or the model being under- and over-predicted and thus, respectively, negative or positive bias (measured by RCE).

The households are grouped according to their RCE_{model} in the intervals: $[-1, -0.3]$, $(-0.3, -0.1]$, $(-0.1, 0.1)$, $[0.1, 0.3)$ and $[0.3, \infty)$. The counts in each interval are plotted in Figure A.10 as a histogram for each quarter, where in each interval the colours indicate the grouping of $RCE_{\text{benchmark}}$ in similar intervals. As an exam-

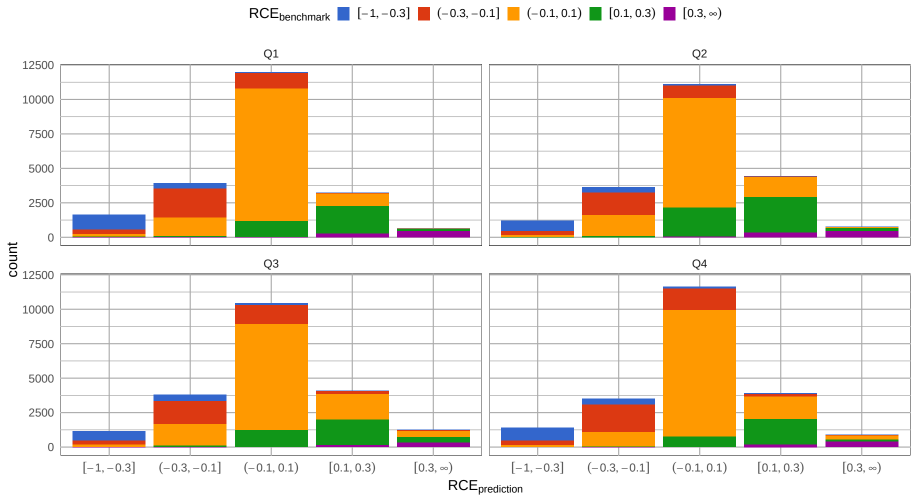


Figure A.10: The RCE of the model predictions distributed according to intervals. For each interval, the distribution of benchmark RCE grouped in similar intervals is indicated by colours. Each plot shows the distributions for the quarter specified above the plot.

ple, it can be seen that, in all four quarters, around half of the households have model predictions where $RCE_{\text{model}} \in (-0.1, 0.1)$, and most of them have also $RCE_{\text{benchmark}} \in (-0.1, 0.1)$ as marked with orange - hence this group have both quite fine model and benchmark predictions (in terms of bias as measured by cumulated error). Another example is the group marked with orange in the left-most group $RCE_{\text{model}} \in [-1, -0.3]$. For them the model was under-predicting, but the benchmark was quite fine.

The histograms show that, for most cases, the model and benchmark RCE fall into the same intervals, although some are shifted, but mostly into the neighbouring interval. The distributions seem to be quite the same in all four quarters, but of course with some variation.

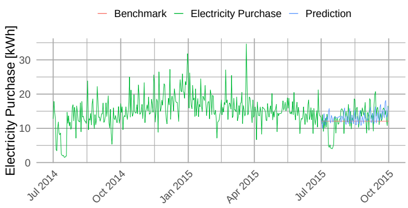
In order to get some idea about typical consumption patterns posing challenges to both the model and the benchmark, households in the different groups are picked out and plotted. First, four examples from the interval where the model yielded good accumulated predictions, $RCE_{\text{model}} \in (-0.1, 0.1)$, are analysed:

- Figure A.11a: In this example the benchmark is also a fine prediction.

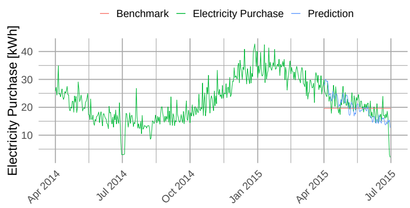
- Figure A.11b: The benchmark is also fine in terms of the accumulated prediction. It can be seen that the residents are absent from the household for one or two weeks due to their vacation, but otherwise they keep their behaviour.

- Figure A.11c: An example where the benchmark over-predicts, $RCE_{\text{benchmark}} \in [0.3, \infty)$. A huge decrease in consumption since last year can be seen. The change occurred some months prior to the prediction period and was caught by the model. The change could be a change in heating system (from electrical to non-electrical), or a lowering in the impact of the temperature difference (better insulation).

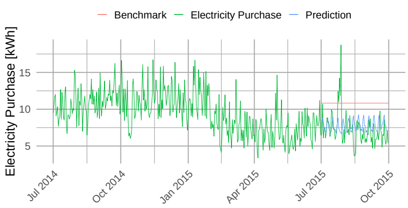
- Figure A.11d: An example where the benchmark under-predicts, $RCE_{\text{benchmark}} \in [-1, -0.3]$. It can be seen that a shift to a higher level of consumption occurred some months prior to the prediction period. Again, the model catches this change, but the benchmark does not.



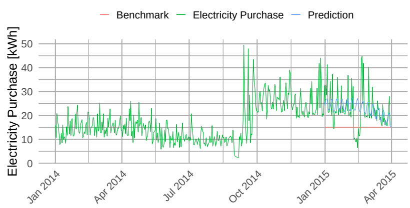
(a)



(b)



(c)



(d)

Figure A.11: Examples where the accumulated model predictions are quite accurate, i.e. $RCE_{\text{model}} \in (-0.1, 0.1)$.

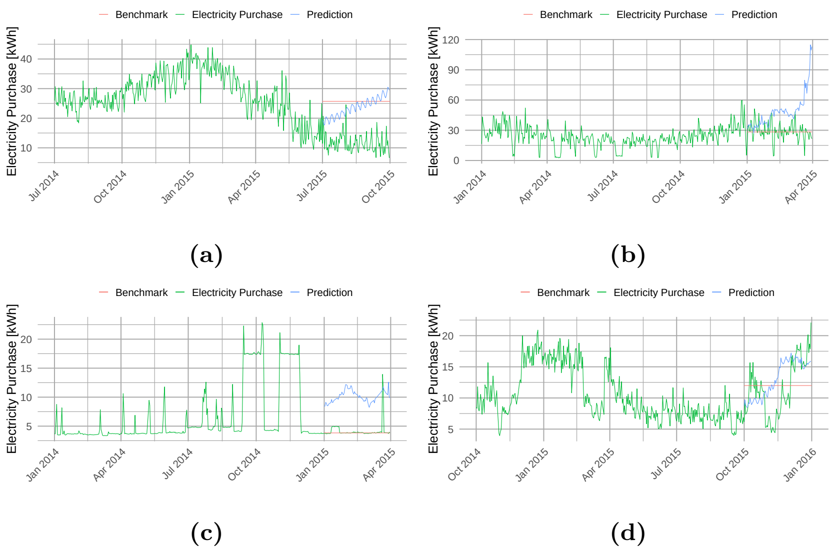
Four examples where the model over-predicts (RCE_{model} in $[0.1, 0.3)$ and $[0.3, \infty)$) have been picked out:

- Figure A.12a: It can be seen that a change in consumption close to the prediction period, around May 2015, is not caught by the model.

- Figure A.12b: A deficiency of the model can be seen, where the impact from temperature explodes on longer horizons. This can be categorized as a kind of Type I error, since data by chance fell out such that the impact was highly over-estimated.

- Figure A.12c: An irregular consumption pattern, most likely a holiday house as the pattern was the same in the quarter one year earlier the pattern was the same and the house was not used in the winter. In this case the benchmark is actually a good prediction.

- Figure A.12d: A household where the consumption shifts at irregular times, and neither the benchmark nor the model can describe this behaviour.



(a) (b) (c) (d)

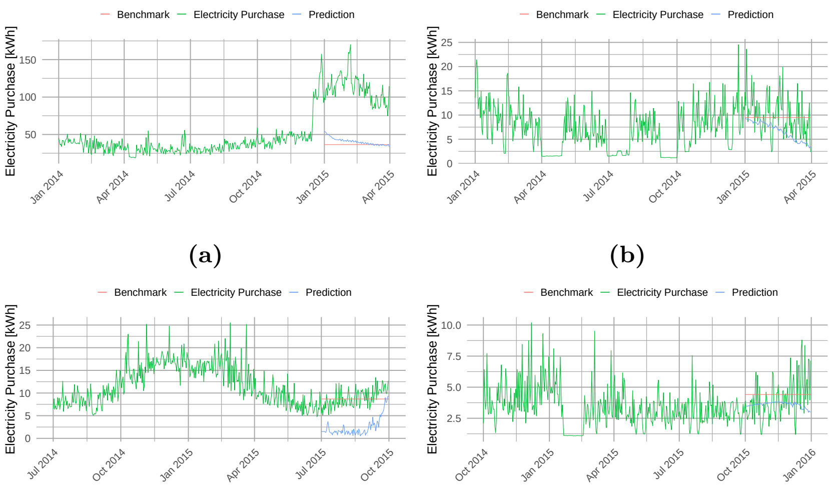
Figure A.12: Examples where the model over-predicts (RCE_{model} in $[0.1, 0.3)$ and in $[0.3, \infty]$).

Finally, examples where the model under-predicts ($\text{RCE}_{\text{model}}$ in $[-1, -0.3]$ and $(-0.3, -0.1])$ have been selected:

- Figure A.13a: Again an example of a late shift before the prediction period which is not captured by the model.

- Figure A.13b: Another example where the residents are absent for periods. The longer horizons are affected by this, leading to a decreasing level during the prediction period.

- Figure A.13c: It is not simple to explain the very low model prediction during the summer quarter, but such outlier examples occur.
- Figure A.13d: The model did not catch the impact of temperature, which kicks in at the end of the prediction period. It is likely that this is caused by high variation during last winter or the low consumption period around Feb. 2014.



(a) (b) (c) (d)

Figure A.13: Examples where the model under-predicts (RCE_{model} in $[-1, -0.3]$ and in $(-0.3, -0.1]$).

It should be noted that the examples presented do not represent all possible consumption patterns, but have been chosen to represent interesting examples. Hence, there are many other reasons for prediction deviation than described in this paper.

A.5 Discussion and Conclusion

This article presents a model procedure for long-term baseline prediction, and results from applying it on daily values from 22,000 households' electricity consumption. The results indicate that, in general, it is possible to improve the predictions in comparison to a one-year-past average benchmark, but also that there is a potential for improvement of the procedure.

From the results obtained, it seems worthwhile to use the presented model procedure in a monitoring tool, which households can use to pinpoint when deviation from the expected baseline occurs. This feedback will be useful for consumers, both as a warning if they slowly start using more electricity and if something out of the ordinary leads to changes in their consumption.

It is possible to improve the model procedure on several points. Firstly, to avoid negative predictions, it is suggested to transform the data using a Box-Cox transformation choosing the transformation parameter such that the transformed data leads to a more symmetric distribution, which when transformed back cannot be negative. A drawback in this is that the parameter coefficients can be difficult to interpret in relation to the electricity consumption.

Secondly, the tuning of parameters for each horizon was as inspired by [17], where the tuning works out very well for short-term forecasting. However, for long-term prediction (i.e. where the re-analysis weather data is used and hence the function from input to output is the same for all horizons), it could be better to use only the one-step-ahead parameters for all horizons.

Thirdly, the effect of using additional inputs, especially the solar radiation and wind effects, should be in investigated further. Also, the tuning of the low-pass filter smoothing factor could be carried out for each individual household.

Finally, it was found that, in the presented model procedure, the changes in consumption 1-3 months back did not influence the predictions and changes 3-5 months back only had a small influence due to high memory/low forgetting. Optimization of the forgetting factor for each household could lead to improvements and this was tried, but resulted in a considerable increase in computation time and some control difficulties due to covariance wind up. To solve these issues, it is suggested to investigate self-tuning control [21], where upper and lower bounds on the co-variance can be implemented to avoid a wind up and a decay to zero and thus retain alertness in the model. With self-tuning control, the tuning of the forgetting factor is done online and hence no optimization of this is needed. The evolution of the parameter coefficients might be used as a crude disaggregation tool to explain the origin of the consumption, e.g. how much of the consumption is due to heating of the household and if it has changed.

Acknowledgements

This work was partly funded by Innovation Fund Denmark (IFD) under File No. 5189-00045B.

References A

[1] *What is DataHub?* `https://en.energinet.dk:443/Electricity/DataHub`. Accessed:2018-02-13. 2018.

[2] C. Fischer. "Feedback on household electricity consumption: a tool for saving energy?" In: *Energy efficiency* 1.1 (2008), pp. 79–104.

[3] P. W. Schultz et al. "Using in-home displays to provide smart meter feedback about household electricity consumption: A randomized control trial comparing kilowatts, cost, and social norms". In: *Energy* 90 (2015), pp. 351–358.

[4] F. M. Andersen, H. V. Larsen, and T. K. Boomsma. "Long-term forecasting of hourly electricity load: Identification of consumption profiles and segmentation of customers". In: *Energy conversion and Management* 68 (2013), pp. 244–252.

[5] F. L. Quilumba et al. "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities". In: *IEEE Transactions on Smart Grid* 6.2 (2015), pp. 911–918.

[6] K. Gajowniczek and T. Zabkowski. "Short term electricity forecasting using individual smart meter data". In: *Procedia Computer Science* 35 (2014), pp. 589–597.

[7] P. Lusis et al. "Short-term residential load forecasting: Impact of calendar effects and forecast granularity". In: *Applied energy* 205 (2017), pp. 654–669.

[8] W. Kong et al. "Short-term residential load forecasting based on resident behaviour learning". In: *IEEE Transactions on Power Systems* 33.1 (2018), pp. 1087–1088.

[9] K. Amasyali and N. M. El-Gohary. "A review of data-driven building energy consumption prediction studies". In: *Renewable and Sustainable Energy Reviews* 81 (2018), pp. 1192–1205.

[10] R. K. Jain et al. "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy". In: *Applied Energy* 123 (2014), pp. 168–178.

[11] F. Lai, F. Magoules, and F. Lherminier. "Vapnik's learning theory applied to energy consumption forecasts in residential buildings". In: *International Journal of Computer Mathematics* 85.10 (2008), pp. 1563–1588.

[12] Y. Iwafune et al. "Short-term forecasting of residential building load for distributed energy management". In: *2014 IEEE International Energy Conference (ENERGYCON)*. IEEE. 2014, pp. 1197–1204.

[13] S. Singh and A. Yassine. “Big data mining of energy time series for behavioral analytics and energy consumption forecasting”. In: *Energies* 11.2 (2018), p. 452.

[14] N. Fumo and M. R. Biswas. “Regression analysis for prediction of residential energy consumption”. In: *Renewable and sustainable energy reviews* 47 (2015), pp. 332–343.

[15] E. N. Lorenz. “A study of the predictability of a 28-variable atmospheric model”. In: *Tellus* 17.3 (1965), pp. 321–333.

[16] D. P. Dee et al. “The ERA-Interim reanalysis: Configuration and performance of the data assimilation system”. In: *Quarterly Journal of the royal meteorological society* 137.656 (2011), pp. 553–597.

[17] P. Bacher et al. “Short-term heat load forecasting for single family houses”. In: *Energy and buildings* 65 (2013), pp. 101–112.

[18] H. A. Nielsen and H. Madsen. “Modelling the heat consumption in district heating systems using a grey-box approach”. In: *Energy & Buildings* 38.1 (2006), pp. 63–71.

[19] H. Madsen. *Time series analysis*. CRC Press, 2007.

[20] P. Vogler-Finck, P. Bacher, and H. Madsen. “Online short-term forecast of greenhouse heat load using a weather forecast service”. In: *Applied Energy* 205 (2017), pp. 1298–1310.

[21] S. L. Shah and W. R. Cluett. “Recursive least squares based estimation schemes for self-tuning control”. In: *The Canadian Journal of Chemical Engineering* 69.1 (1991), pp. 89–96.

Paper B

Extensions to the Recursive Least Squares modelling scheme. Self-tuning control and Box-Cox transformation

Authors:
Jon Liisberg, Jan Møller, Peder Bacher

Technical report:
DTU library.

Extensions to the Recursive Least Squares modelling scheme. Self-tuning control and Box-Cox transformation

Jon Liisberg[1 2], Jan Møller[1], Peder Bacher[1]

Abstract

Self-tuning control (STC) is implemented to a Recursive Least Squares (RLS) modelling scheme for daily long-term prediction of electricity consumption. Also, 1-parameter Box-Cox transformation of the electricity consumption is implemented. With these the STC extension to the modelling scheme it is tested if adaptation is faster and if alertness is retained in the model. With the transformation is it investigated of this will solve the issue with negative prediction observed for the RLS.

It is found that in some cases the extension show great improvement in comparison to the ordinary RLS. Using parameter coefficients for the one-day ahead horizon in all horizons of the prediction period, showed that an improvement can be achieved in some cases.

B.1 Introduction

Working with time-series on electricity consumption in residences, we have observed huge deviation in the consumption patterns. Furthermore, we have observed consumer patterns to change over time, suddenly change (shock), or suddenly change due to exogenous variables (Turning electrical heating on/off).

In this report the focus is to test whether introducing self-tuning control to the ordinary Recursive Least Squares (RLS) with forgetting can improve prediction, speed up adaptation and retain alertness for these types of time-series. This is tested on two selected time-series, one to illustrate the alertness retention and one where shock was observed.

[1]DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark
[2]SEAS-NVE, Digital Innovation, Hovedgaden 36, DK-4520 Svinninge, Denmark

Furthermore, box-cox transformation is introduced and applied on a household where the ordinary RLS yielded negative values in the forecast.

For both cases only using the 1-day a-head model parameters to do daily predictions for the whole quarter is investigated.

This report is meant as an extension to the model scheme presented in [1](Paper A), hence it is highly recommended to read this prior to the present report.

B.2 Method

B.2.1 Recursive least squares with forgetting

RLS with forgetting is a parametric model where the implementation is based on [2]. To briefly give an idea of the RLS algorithm it is shown for horizon k below:

Update step:

$$\mathbf{R}_t = \lambda \mathbf{R}_{t-1} + \mathbf{X}_t \mathbf{X}_t^T \tag{B.1}$$

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \mathbf{R}_t^{-1} \mathbf{X}_t (Y_t - \mathbf{X}_t^T \hat{\boldsymbol{\theta}}_{t-1}) \tag{B.2}$$

Prediction step

$$\hat{Y}_{t+k} = \mathbf{X}_{t+k}^T \hat{\boldsymbol{\theta}}_t \tag{B.3}$$

Here $\hat{\boldsymbol{\theta}}_t$ is the parameter vector, Y_t is the observation, $\hat{Y}_t$ is the prediction and $\mathbf{X}_t$ is the vector of variables at time t, $\mathbf{R}_t$ is the inverse sample covariance matrix and λ is the forgetting factor. One model for each horizon k is fitted, hence for a quarterly forecast we need 90-92 models.

B.2.2 Model selection

The model selection scheme for the current set up is based on two stages. The first is a test of overall dependence on exogenous variables using forward selection based on linear regression of the output and input variables. The input variables consist of:

- Temperature difference between a fixed indoor temperature and the outdoor temperature.

- Low pass filtered temperature difference to model the envelope of the house.
- The length of the night given location of the household.

The second stage is for each horizon where harmonics is chosen based on optimization using root mean squared error (RMSE) as the objective function.

$$RMSE = \sqrt{\frac{\sum_{t=k}^{n}(Y_t - \hat{Y}_t)^2}{n-k}} \tag{B.4}$$

The model selection scheme is described in detail in [1](Paper A)

B.2.3 Recursive least squares with variable forgetting and covariance resetting

Based on [2] and [3] a method with variable forgetting and covariance resetting in the RLS scheme has been implemented. To set up RLS with variable forgetting we first need to restructure the update step in the RLS algorithm (equations B.1 and B.2) to:

$$\hat{\boldsymbol{\theta}}_t = \hat{\boldsymbol{\theta}}_{t-1} + \mathbf{K}_t(Y_t - \mathbf{X}_t^T \hat{\boldsymbol{\theta}}_{t-1}) \tag{B.5}$$

$$\mathbf{K}_t = \frac{\mathbf{P}_{t-1}\mathbf{X}_t}{\lambda + \mathbf{X}_t^T \mathbf{P}_{t-1}\mathbf{X}_t} \tag{B.6}$$

$$\mathbf{P}_t = 1/\lambda \left(\mathbf{P}_{t-1} - \mathbf{K}_t \mathbf{X}_t^T \mathbf{P}_{t-1}\right). \tag{B.7}$$

Here $\mathbf{K}_t$ is the gain vector and $\mathbf{P}_t = \mathbf{R}_t^{-1}$ is the co-variance matrix. The implementation of variable forgetting factor is done by replacing Equations (B.6) and (B.7) with the following:

$$\mathbf{K}_t = \frac{\mathbf{P}_{t-1}\mathbf{X}_t}{1 + \mathbf{X}_t^T \mathbf{P}_{t-1}\mathbf{X}_t} \tag{B.8}$$

$$\lambda_t = 1 - \frac{(Y_t - \mathbf{X}_t^T \hat{\boldsymbol{\theta}}_{t-1})^2}{\sigma(1 + \mathbf{X}_t^T \mathbf{P}_{t-1}\mathbf{X}_t)} \tag{B.9}$$

$$\mathbf{W}_t = \mathbf{P}_{t-1} - \mathbf{K}_t \mathbf{X}_t^T \mathbf{P}_{t-1} \tag{B.10}$$

if trace of $\mathbf{W}_t/\lambda_t \leq C$

$$\overline{\mathbf{P}}_t = \mathbf{W}_t/\lambda_t \tag{B.11}$$

else

$$\overline{\mathbf{P}}_t = \mathbf{W}_t. \tag{B.12}$$

Here C is a constant that ensures an upper-bound on the co-variance matrix avoiding covariance wind-up. σ is defined as $\sigma/\sigma_w = 1000$ [3] where σ_w is chosen to be the standard deviation of the residuals.

$$\sigma_w(t) = \sqrt{\frac{\sum_{i=k}^{t}(Y_i - \hat{Y}_i)^2}{(n-k)(1+\mathbf{X}_t^T\mathbf{P}_{t-1}\mathbf{X}_t)}}. \tag{B.13}$$

Equation (B.13) is a global linear estimate of the standard deviation. For horizon k we do not have an estimate of $\hat{Y}_i$ when $i < k$, hence we divide with $N - k$. This is also used as the standard deviation in calculation of prediction intervals.

Let $\overline{\mathbf{P}}_t$ be defined as in Equation (B.11) or (B.12), if the trace of $\overline{\mathbf{P}}_t < c_{min}$, c_{min} is a lower bound, then:

$$\mathbf{P}_t = \overline{\mathbf{P}}_t + \mathbf{Q} \tag{B.14}$$

where $\mathbf{Q} = c\mathbf{I}$. for a suitable scalar c and where $\mathbf{I}$ is the identity matrix, else:

$$\mathbf{P}_t = \overline{\mathbf{P}}_t. \tag{B.15}$$

This will ensure a lower bound on the covariance matrix and stopping it from decaying to zero. This lower bound is controlled by the choice of c_{min} and c.

For both the ordinary and the self-tuning RLS the parameters was initialized as in Table B.1.

Table B.1: Tuning parameter initialization

Parameter	value	description
P_0	10000$\mathbf{I}$	a diagonal matrix
C	10000	upper bound of P
c_{min}	1	lower bound of P
c	0.01	
λ_{min}	0.5	
λ_{RLS}	0.999	

The chosen parameters are found by testing many different consumption patterns until robust result for all patterns was obtained.

B.2.4 Box-Cox transformation of data

Since $y \in [0, \infty)$ the data might be transformed when there is large variation in the data to better work with normal distributions, we use the 1-parameter Box-Cox transformation

$$y_i^{(\gamma)} = \begin{cases} \frac{y_i^\gamma - 1}{\gamma} & \gamma \neq 0 \\ log(y_i) & \gamma = 0 \end{cases} \tag{B.16}$$

Here γ is optimized such that $\mathbf{y}^{(\gamma)}$ resembles a normal distribution the best for $\gamma \in [-2, 4]$. The chosen interval for γ is based on Figure B.1 and issues with the inverse transformation for $\gamma < -2$. It should be noted that the when the prediction is transformed back it is not an estimate of the mean but only the median.

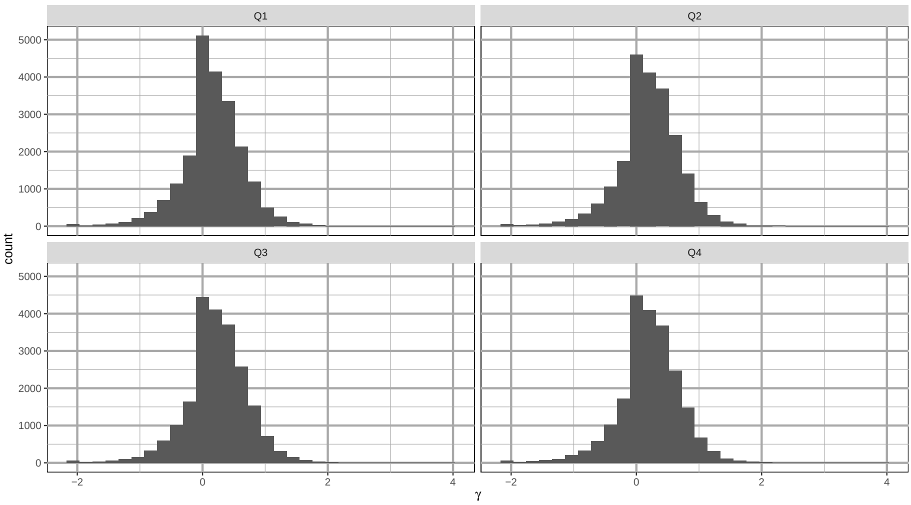


Figure B.1: The distribution of transformation coefficient tested on 22000 time-series

B.3 Results

B.3.1 STC and resetting

To illustrate how self-tuning control in RLS is working an example is presented in Figure B.2. This example has high variation between days, but the yearly variation is not changing much, hence no temperature dependence was expected in this example. However, temperature was found to enter the model with a high threshold hence there is input from temperature in almost all the training steps.

In Figure B.2 four different prediction models are shown, two with high memory and two with low memory. The high memory models are the RLS and using the average from the last year in the same period. From Table B.2 it is clear that the high memory models are the best for this particular time-series. The

RLS learn over time that the temperature parameter does not influence the consumption significant and it does not affect the prediction much. The STC models are highly influenced by the last training day in the model fitting (the longer horizons are not trained on the most recent data prior to prediction) and are varying more than the high memory RLS.

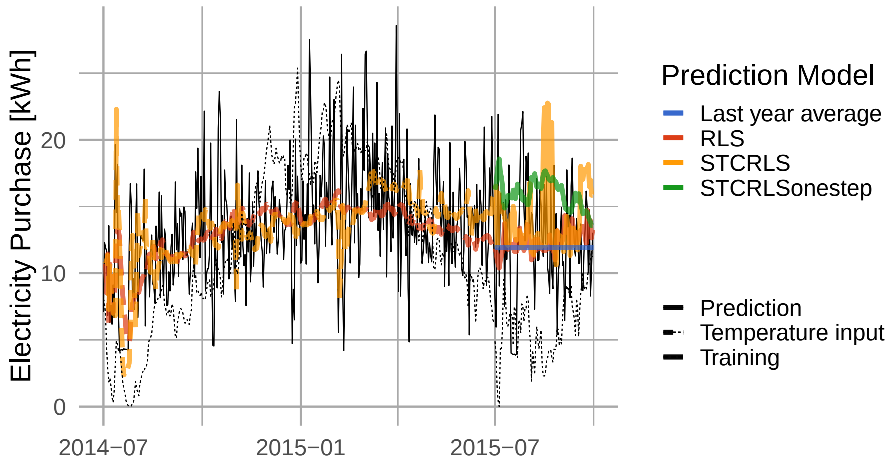


Figure B.2: 1 step ahead model fit of the ordinary and self-tuning RLS

Table B.2: Relative cumulated error (RCE) and sum of squared error (SSE) for each model type

Model type	RCE	SSE
RLS	-0.01	1703.2
STCRLS	0.13	2711.1
STCRLSonestep	0.27	2766.7
Last year average	-0.04	1639.5

The example in Figure B.2 is not shown for the performance in the model fitting, but because the evolution of the covariance matrix P shows how the alertness is retained while training the model. In Figure B.3 the evolution of the fitted values $(\hat{Y})$, the trace of the covariance matrix P, the global linear estimate of σ_w and the forgetting factor(memory) λ are shown for both the ordinary RLS and the self-tuning RLS for prediction horizon $k = 1$.

In Figure B.3 the trace of the covariance matrix P quickly decays toward zero for the ordinary RLS. This can ensure a stable evolution of the parameter coefficients, but sometimes it can become to small such that the gain will also decay

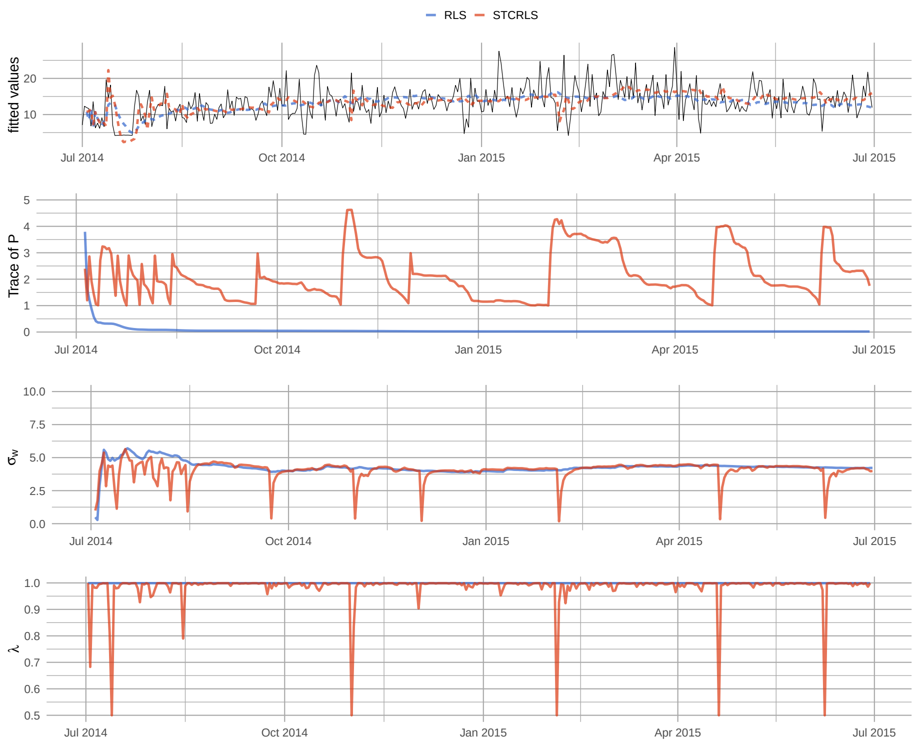


Figure B.3: The evolution of parameters controlling memory and alertness for horizon $k = 1$

towards zero and then errors will not update the parameter coefficients and the model has lost its alertness.

For the STC model the trace of P is varying a lot more and the lower bound specified in Table B.1 is clearly seen. There are mainly two reasons for the trace to jump up from this bound. Either the lower bound is hit and the c is added to the diagonal of P, or the error between fitted value and the observed consumption is big which results in a small lambda value. The latter is the main reason in this example but here also the temperature input could influence the evolution of P when there are big jumps in the temperature difference.

For the global linear estimation of the standard deviation it is observed that they are similar for both the RLS and STC. For the STC it quickly adapts back after resetting.

B.3.2 Household with several jumps in consumption level

To illustrate a situation in which self-tuning control RLS could be beneficial, an example of this is shown in Figure B.4. For this time-series there are several jumps in the consumption over time. The high memory RLS adapts slowly to the changes while the STC quickly adapts to the new level.

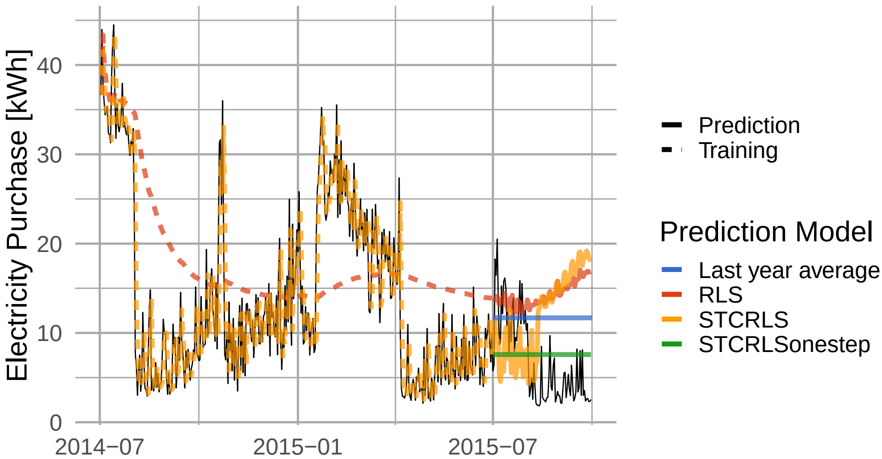


Figure B.4: 1 step ahead model fit of the ordinary and self-tuning RLS

Looking into the prediction performance of the models a clear improvement is seen using the STCRLS model. But the long horizons of both models still depend highly less recent data. If only using the parameter coefficients for $k = 1$ the prediction can rely on the most recent data observed (The model for $k = 1$ is a model only with intercept hence the same prediction in each horizon). In this case it was the better choice but next quarter a different model could be better if the consumption jumps again.

Table B.3: Relative cumulated error (RCE) and sum of squared error (SSE) for each model type

Model type	RCE	SSE
RLS	1.09	7738.9
STCRLS	0.76	8720.7
STCRLSonestep	0.11	2090.5
Last year average	0.73	4213.6

From Figure B.5 it is seen that the trace of P is much more stable in comparison to the previous example, here we only see spikes when there are jumps in the consumption. The estimates of σ_w are very different between the RLS and the STCRLS, For the RLS some information from jump to jump seem to be kept where for the STCRLS it is only between jump variation that is remembered.

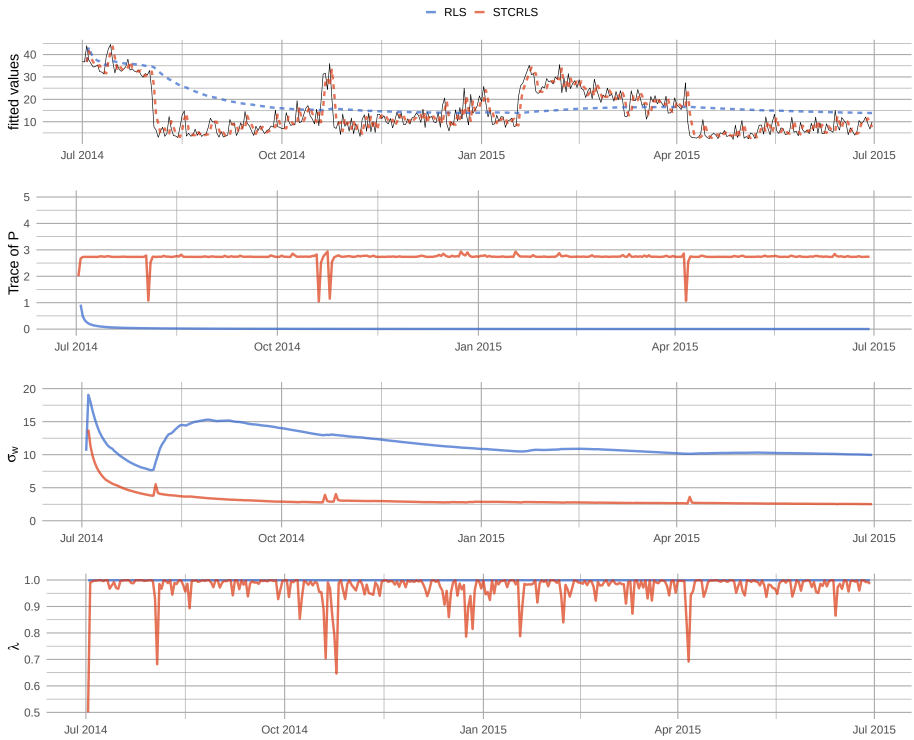


Figure B.5: The evolution of parameters controlling memory and alertness for horizon $k = 1$

B.3.3 Ordinary RLS with negative predictions

To avoid negative forecast Box-Cox transformation is introduced as an extension to the RLS modelling scheme. In Figure B.6 an example on the ordinary RLS yielding negative predictions is shown. The issue for the RLS is that the intercept has become negative during the training for the short horizons, but for the long horizons the intercept is positive as if it was in the middle of winter and we get very high predictions. By transforming the data and fitting the RLS on this data we get the RLSBC shown in Figure B.6. Here we still have the issue with high values of intercept in the long horizons. instead the parameter coefficients for $k = 1$ are used for predicting all horizons here using the RLSBC parameter coefficients, shown in Figure B.6 as RLSBConsetep.

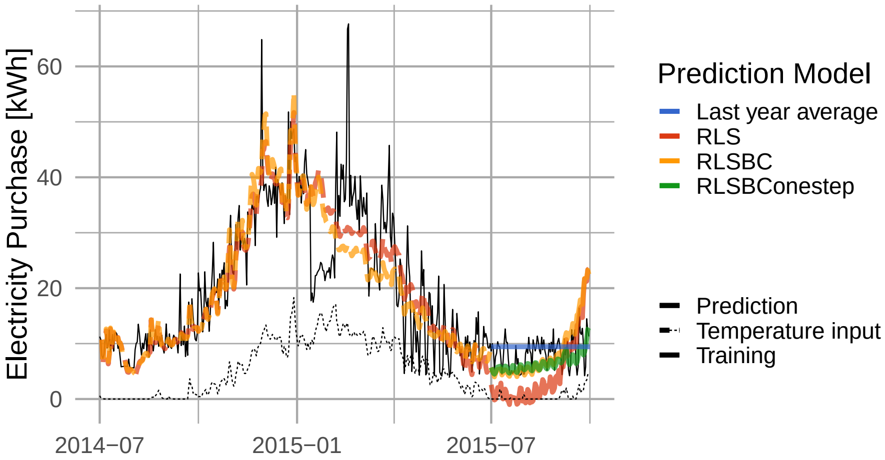


Figure B.6

Looking into the prediction performance in Table B.4 the RLSBC has a very good RCE but this is coincidental since due to having low predictions in the beginning of the quarter and high in the end, hence it has the worst SSE. The RLSBConsetep performs better, but there seem to have been a change in the day to day variation from last year which is not captured.

Table B.4: Relative cumulated error (RCE) and sum of squared error (SSE) for each model type

Model type	RCE	SSE
RLS	NA	4587.8
RLSBC	-0.01	2192.2
RLSBConestep	-0.23	886.6
Last year average	0.12	778.8

B.4 Discussion/Conclusion

With these extensions and the option to choose parameters from shorter horizon models to predict the long horizons, we have a wide range of models capable of handling many different consumption patterns and changes within these. It was shown that alertness is retained in the model using the self-tuning scheme, but also that the high memory models can be more robust in certain settings. By transforming data using Box-Cox transform the issues with negative predictions are resolved but further work is still needed to resolving problem predicting consumption when the heating is turned on or off.

The tuning parameters for the STC has been fixed but it is possible investigate if individual optimization of these is worthwhile or if the computational effort is too high.

Further investigation of which model type are best suited to different consumption patterns are needed and metrics for consumption patterns needs to be developed in order to automate this part of the model selection.

Overall this preliminary study indicates that further investigation of self-tuning in the RLS could be fruitful in the performance of the quarterly forecast of electricity consumption when dealing with the vast variation in consumption patterns.

Acknowledgement

This work is partly funded by Innovation Fund Denmark (IFD) under File No. 5189-00045B.

References B

[1] J. Liisberg, J. Møller, and P. Bacher. "Models for long-term baseline prediction of daily electricity consumption in individual households". In: (). submitted for consideration to Applied Energy.

[2] H. Madsen. *Time series analysis*. CRC Press, 2007.

[3] S. L. Shah and W. R. Cluett. "Recursive least squares based estimation schemes for self-tuning control". In: *The Canadian Journal of Chemical Engineering* 69.1 (1991), pp. 89–96.

Paper C

Hidden Markov Models for indirect classification of occupant behaviour

Authors:
Jon Liisberg, Jan Møller, Hans Bloem, Jordi Cipriano, Gerard Mor, Henrik Madsen

Published in:
Sustainable Cities and Society, Volume 27, November 2016, Pages 83-98 .

Hidden Markov Models for indirect classification of occupant behaviour

Jon Liisberg[1,2], Jan Møller[1], Hans Bloem[3], Jordi Cipriano[4], Gerard Mor[3], Henrik Madsen[1]

Abstract

Even for similar residential buildings, a huge variability in the energy consumption can be observed. This variability is mainly due to the different behaviours of the occupants and this impacts the thermal (temperature setting, window opening, etc.) as well as the electrical (appliances, TV, computer, etc.) consumption.

It is very seldom to find direct observations of occupant presence and behaviour in residential buildings. However, given the increasing use of smart metering, the opportunity and potential for indirect observation and classification of occupants' behaviour is possible. This paper focuses on the use of Hidden Markov Models (HMMs) to create methods for indirect observations and characterisation of occupant behaviour.

By applying homogeneous HMMs on the electricity consumption of fourteen apartments, three states describing the data were found suitable. The most likely sequence of states was determined (global decoding). From reconstruction of the states, dependencies like ambient air temperature were investigated. Combined with an occupant survey, this was used to classify/interpret the states as 1) Absent or asleep, 2) Home, medium consumption and 3) Home, high consumption. From the global decoding, the average probability profiles with respect to time of day were investigated, and four distinct patterns of occupant behaviour were observed. Based on the initial results of the homogeneous HMMs and with the observed dependencies, time dependent HMMs (inhomogeneous HMMs) were developed, which improved forecasting. For both the homogeneous and inhomogeneous HMMs, indications of common parameters were

[1]DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark
[2]SEAS-NVE, Digital Innovation, Hovedgaden 36, DK-4520 Svinninge, Denmark
[3]European Commission, JRC – IET, RE-Unit, TP 450, I-21020 Ispra, Italy
[4]CIMNE-UPC, Terrassa, Spain

observed, which suggests further development of the HMMs as population models.

C.1 Introduction

In building-design optimisation, energy diagnosis, performance evaluation and building energy simulations, the impact of the occupants' behaviour is often under-recognised and over-simplified. The influences of occupant behaviour are complex and stochastic. In recent years, the importance of occupant behaviour has been recognised, and many new approaches have been developed to model the effect of occupant behaviour. To achieve an overview of the approaches in modelling occupant behaviour in buildings, a small literature study has been carried out.

In building-related models for occupant behaviour, there are two main focus areas, (1) occupant presence and movement and (2) occupant interaction with indoor climate (adjusting a thermostat, opening a window for ventilation, turning on lights or closing blinds). Studies related to these areas are typically related to either residential or commercial buildings [1, 2].

The following is a cursory review of the papers in the literature study.

In modelling occupant presence and movement in office buildings, both homogeneous [3] and time-inhomogeneous [4, 1] Markov chains have been used. The models are used as input for building energy simulations. In [1] a comparison of the performance between homogeneous and time-inhomogeneous models was carried out, and the inhomogeneous model was found to be superior.

In the work on modelling overtime schedules in office buildings, [5] uses a binomial distribution to represent the number of occupants working overtime and an exponential distribution to describe the duration of overtime. The overtime model is used to generate overtime schedules as input to building energy simulations.

Based on seven years of measuring window opening and closing behaviour, three modelling methods for prediction of actions on windows were developed [6]. The methods are logistic probability distributions, Markov chains and continuous-time random processes.

In a field study of the thermal comfort of office occupants [7], logistic regression was used to predict the probability of occupants' actions.

In a simulation study of an adaptive automation system for the visual comfort of office occupants [8], the models for predicting occupants' turning light on/off and opening/closing blinds are based on Markovian state transition probabilities.

For air-conditioning in residences, [9] identifies on/off state transition probability functions dependent on indoor and outdoor temperature. These functions are requisite for applying a Markov model to a cooling schedule.

A methodology to predict residential occupants' time-dependent activities is presented in [10]. Using a time-use survey, the model is calibrated based on three time-dependent quantities: (1) the probability of being home, (2) the conditional probability of starting an activity while at home, and (3) the probability distribution function for the duration of the activity. Transitions between activity types are modelled as an inhomogeneous Markov process.

Studies in building energy simulations [11, 2] have investigated the impact of changing from standardised occupant behaviour profiles to a probabilistic approach in simulating these profiles. [2] showed a large increase in energy consumption, with this approach.

Based on data mining using cluster analysis, [12] examines the influences of occupant behaviour on building energy consumption. A methodology for identifying energy-inefficient behaviour in residential buildings was developed.

[13] provides an overview of recent studies undertaking predictive and descriptive tasks in the building field. This is done by using data-mining techniques to extract hidden but useful knowledge. For occupant behaviour, a key issue is to understand the interactions between occupant behaviour and other influencing factors.

From this literature study, different approaches seem highly problem-specific. Many use Markov chains/processes in the description of the transition between presence, non-presence, movement between rooms and transitions between activities. This indicates that Markov chains/processes are highly useful for modelling occupant behaviour in a wide range of settings. With the idea to extract hidden knowledge from data, and using Markov chains to model occupant behaviour, this has spurred us to look at methods to observe occupant behaviour in an indirect manner e. g. Hidden Markov Models (HMMs).

When measuring the electricity consumption in similar residential buildings, the variability in the consumption is often very large. This is mainly due to the diversity of occupant behaviour. The occupants not only impact the electricity consumption, but also the general energy consumption [14, 15, 16]. Due to privacy concerns, and the cumbersome work of obtaining direct observations of occupant behaviour, indirect means of classifying occupant behaviour are needed. Several models have been developed for simulation purposes using data-mining approaches [15]. Based on a time-use survey, it is suggested that occupant behaviour in residential buildings could be classified according to the following three states: (1) at home and awake, (2) sleeping, or (3) absent [14]. Given the increasing use of smart metering by the utilities, the potential of using these metering data for indirect classification of residential occupant behaviour is now possible. Applying a homogeneous Hidden Markov Model (HMM) to electricity consumption data from a residence results in a number of states that could be interpreted in a similar manner [17].

The focus of the study presented in this paper is to investigate the applications of HMMs on frequent observations of electricity consumption in residences. The study seeks to test the hypothesis that, by applying HMMs on observations of electricity consumption, we can:

1. Classify the states of the HMM, i. e. of the occupant(s) in accordance to occupant behaviour.
2. Identify possible covariates/explanatory variables.
3. Forecast and simulate future energy consumption.

1), 2) and 3) can be solved by both homogeneous and time-inhomogeneous models. It is suggested that to improve the capabilities for forecasting and simulation, covariates/explanatory variables and time-inhomogeneous Markov chains, are needed [1].

The study also seeks to investigate whether the HMMs for each residence can be collected in population models [18] to forecast or simulate groups of residences.

The aim is to present a modelling framework for HMMs on frequent observations of electricity consumption, and then apply this framework to several residential apartments. Focus will be on interpreting the states of the HMMs to validate the models and suggest further development of these models.

The outcome is an initial framework for using HMMs on frequent observations of electricity consumption and proposals for further model development. This study is a further elaboration of some of the results in [19].

Nomenclature

AIC	Akaike information criterion
BIC	Bayesian information criterion
HMM	Hidden Markov Model
CRPS	continuous rank probability score
cdf	cumulative distribution function
pdf, $p(x)$	probability mass or density function
m	number of states
t, s	a time stamp in discrete time
T	maximum of t, i.e. $t \in \{1, ..., T\}$
$\mathbb{N}$	the natural numbers
$\mathbb{R}$	the real numbers
$i, j, k \in \mathbb{Z}$	integers
C_t	the state of a Markov chain at time t
X_t	the state of the random process $\{X_t\}$ at time t
x_t	the observation of the random process $\{X_t\}$ at time t
$\boldsymbol{A}, \boldsymbol{\Gamma}$	matrices
$\boldsymbol{a}, \boldsymbol{\theta}$	row vectors
$\boldsymbol{a'}$	a column vector

C.2 Methods

This section contains a brief introduction to Hidden Markov Models and a description of the methods used in the implementation and validation of the Hidden Markov Models.

C.2.1 Hidden Markov Models

A Hidden Markov Model (HMM) consists of two components; an independent mixture model and a Markov chain.

An independent mixture distribution consists of a finite number of component distributions, and a mixing distribution. The component distributions can be either discrete or continuous. For m components, the mixture distribution depends on m probability or density functions

$$\text{component distribution } \delta_1, \cdots, \delta_m$$
$$\text{probability or density functions } p_1(x), \cdots, p_m(x)$$

The component is specified by the discrete random variable C which performs the mixing where $Pr(C = i) = \delta_i$ for $i \in \{1, \cdots, m\}$ and $\sum_{i=1}^{m} \delta_i = 1$. Let X denote the random variable which has mixture distribution. Then the probability or density function of X is given by:

$$p(x) = \sum_{i=1}^{m} \delta_i p_i(x). \tag{C.1}$$

The second building block of HMMs is Markov chains. A sequence of discrete random variables $\{C_t : t \in \mathbb{N}\}$ is a discrete-time Markov chain if, for all $t \in \mathbb{N}$, the Markov property is satisfied, i.e.

$$Pr(C_{t+1}|C_t, ..., C_1) = Pr(C_{t+1}|C_t). \tag{C.2}$$

The conditional probabilities, $Pr(C_{s+k} = j|C_s = i)$, called transition probabilities, are the probabilities of $C = j$ at time $s + k$ given $C = i$ at time s. If the transition probabilities do not depend on time, then the chain is called homogeneous, otherwise inhomogeneous. The k-step transition probability for a homogeneous Markov chain is denoted as:

$$\gamma_{ij}(k) = Pr(C_{s+k} = j|C_s = i). \tag{C.3}$$

In particular, $\gamma_{ij}(1)$ is denoted γ_{ij} and can be collected in the transition probability matrix $\mathbf{\Gamma}$.

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix}. \tag{C.4}$$

Further, it can be shown, (see e.g. [20]) that for a homogeneous Markov chain $\mathbf{\Gamma}(k) = \mathbf{\Gamma}^k$.

Let $\mathbf{X}^{(T)}$ denote $(X_1, \cdots, X_T)$ and $\mathbf{C}^{(T)}$ denote $(C_1, \cdots, C_T)$. Collecting both parts, a first order HMM can be summarized by:

$$Pr(C_t|\mathbf{C}^{(T-1)}) = Pr(C_t|C_{t-1}), t = 2, 3, ... \tag{C.5}$$

$$Pr(X_t|\mathbf{X}^{(T-1)}, \mathbf{C}^{(T)}) = Pr(X_t|C_t), t \in \mathbb{N} \tag{C.6}$$

Hence, the dynamics is described by the unobserved parameter process $\{C_t : t = 1, 2, ...\}$, which describes the evolution of the states in time. The observations are described by the state-dependent process $\{X_t : t = 1, 2, ...\}$ such that when C_t is known, the distribution of X_t only depends on the current state C_t. The structure of a first-order Hidden Markov Model is illustrated in Figure C.1.

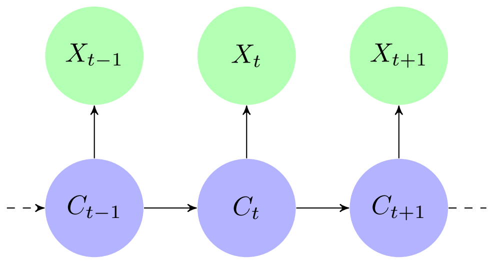


Figure C.1: Directed graph of a Hidden Markov Model.

C.2.1.1 Global decoding and other conditional probabilities for HMMs

Given a HMM and observations, information can be deduced about the states occupied by the underlying Markov chain. The key inferential tools are conditional probabilities, and we will consider some of them in the following.

Let $\mathbf{X}^{(-t)}$ denote $(X_1, \cdots, X_{t-1}, X_{t+1} \cdots, X_T)$, then the conditional distributions of X_t given all other observations are given by:

$$Pr(X_t = x|\mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}). \tag{C.7}$$

This is used for calculating pseudo-residuals to validate the HMMs, as described later in Section C.2.3.3.

Forecasting distributions are used for the likelihood estimation, and are given by:

$$Pr(X_{t+k} = x|\mathbf{X}^{(t)} = \mathbf{x}^{(t)}). \tag{C.8}$$

This can be used for forecasting in general. However, for maximum likelihood estimation of the model parameters we must use $k = 1$.

Global Decoding is the determination of the most likely sequence of states conditioned on the observations. This sequence is obtained by maximizing the conditional probability:

$$Pr(\mathbf{C}^{(T)} = \mathbf{c}^{(T)} | \mathbf{X}^{(T)} = \mathbf{x}^{(T)}). \tag{C.9}$$

Loosely speaking, Equation (C.9) is the probability of $\mathbf{c}^{(T)}$ given all observations and model. Hence $\mathbf{c}^{(T)}$ can be used to identify model deficiencies.

C.2.2 Parameter estimation

The likelihood principle is used to estimate the parameters of the models. The likelihood function is denoted $L(\boldsymbol{\theta}, \mathbf{x}^{(T)})$, where $\boldsymbol{\theta}$ is the parameters and $\mathbf{x}^{(T)} = (x_1, ..., x_T)$ is the observations. The maximum likelihood parameter estimates are found as:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \mathbf{x}^{(T)}) \tag{C.10}$$

Using the Markov property, this reduces to

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \prod_{t=1}^{T} Pr(X_t = x_t | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}). \tag{C.11}$$

In particular, the discrete likelihood [17] is used for the parameter estimation in this paper, which is calculated as:

$$\begin{aligned} L(\boldsymbol{\theta}, \mathbf{x}^{(T)}) &= \prod_{t=1}^{T} Pr(X_t = x_t | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}) && \text{(C.12)} \\ &= \prod_{t=1}^{T} Pr(X_t = x_t | C_t = i) Pr(C_t | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}) && \text{(C.13)} \\ &= \prod_{t=1}^{T} \sum_{i=1}^{m} \delta_i \cdot Pr(X_t = x_t | C_t = i). && \text{(C.14)} \end{aligned}$$

To avoid unbounded values likelihood of the mixture, we transform density to probability by:

$$Pr(X_t = x_t | C_t = i) = \int_{a_t}^{b_t} p_i(x, \boldsymbol{\theta}_i) dx, \tag{C.15}$$

with $a_t = x_t - \Delta, b_t = x_t + \Delta$ for suitable Δ, and we have

$$L(\boldsymbol{\theta}, \mathbf{x}^{(T)}) = \prod_{t=1}^{T} \sum_{i=1}^{m} \delta_i \int_{a_t}^{b_t} p_i(x_t, \boldsymbol{\theta}) dx. \tag{C.16}$$

C.2.2.1 Estimation routine

The estimation is typically carried out by direct optimization of the likelihood $L(\boldsymbol{\theta}, \mathbf{x}^{(T)})$. In `R`[5] the optimizer used is the `nlm()` optimisation function. This is an unconstrained optimizer, but the optimisation problem is constrained since the parameters for the state-dependent distribution might be constrained. E.g. the gamma distribution has shape k_i and scale θ_i as parameters and these must be greater than zero.

For the transition probability matrix $\boldsymbol{\Gamma}$, the rows must add to one and all parameters γ_{ij} must be non-negative.

For the parameters in the state-dependent distributions, the constraints can be circumvented by an appropriate transformation of the parameters k_i and θ_i. This is done by defining $\eta_{k_i} = \log k_i$ and $\eta_{\theta_i} = \log \theta_i$ for $i = 1, ..., m$, then $\eta_{k_i}, \eta_{\theta_i} \in \mathbb{R}$. After the maximisation of the likelihood with respect to the unconstrained parameters, then the constrained parameters are obtained by transforming back, i.e. $\hat{k}_i = \exp(\hat{\eta}_{k_i})$ and $\hat{\theta}_i = \exp(\hat{\eta}_{\theta_i})$.

For the transition probability matrix $\boldsymbol{\Gamma}$, there are m^2 entries, but only $m(m-1)$ free parameters due to the row-sum constraint. One possible transformation between the m^2 constrained probabilities γ_{ij} and the $m(m-1)$ unconstrained $\tau_{ij} \in \mathbb{R}, i \neq j$ is shown for $m = 3$. Define the matrix:

$$\mathbf{T} = \begin{pmatrix} - & \tau_{12} & \tau_{13} \\ \tau_{21} & - & \tau_{23} \\ \tau_{31} & \tau_{32} & - \end{pmatrix}, \tag{C.17}$$

with $m(m-1)$ entries $\tau_{ij} \in \mathbb{R}$. Let $g : \mathbb{R} \to \mathbb{R}^+$ be a strictly increasing function such as $g(x) = \exp(x)$. Then define

$$\varrho_{ij} = \begin{cases} g(\tau_{ij}) & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}, \tag{C.18}$$

and set $\gamma_{ij} = \varrho_{ij} / \sum_{k=1}^{m} \varrho_{ik}$.

[5] `https://www.r-project.org/`

The parameters γ_{ij}, k_i and θ_i are referred to as the natural parameters [17],while the parameters τ_{ij}, η_{k_i} and η_{θ_i} are referred to as the working parameters. Using these transformations of $\mathbf{\Gamma}, \mathbf{k}$ and $\boldsymbol{\theta}$ the calculation of the maximum-likelihood parameters can be calculated in two steps:

1. Maximise L_T with respect to the parameter $\mathbf{T} = \{\tau_{ij}\}, \boldsymbol{\eta}_k = (\eta_{k_1}, ..., \eta_{k_m})$ and $\boldsymbol{\eta}_\theta = (\eta_{\theta_1}, ..., \eta_{\theta_m})$
2. Transform the estimates of the working parameters to estimates of the natural parameters; i. e.

$$\hat{\mathbf{T}} \rightarrow \hat{\mathbf{\Gamma}}, \hat{\boldsymbol{\eta}_k} \rightarrow \hat{\mathbf{k}}, \hat{\boldsymbol{\eta}_\theta} \rightarrow \hat{\boldsymbol{\theta}}.$$

C.2.2.2 Covariates in the transition probabilities

One way to model time variations and seasonality in HMMs is to drop the assumption of a homogeneous Markov chain and assume that the transition probabilities are functions of time, which leads to inhomogeneous Markov models. For m states, the transition probability will be denoted by

$$_t\mathbf{\Gamma} = \begin{pmatrix} {}_t\gamma_{11} & \cdots & {}_t\gamma_{1m} \\ \vdots & \ddots & \vdots \\ {}_t\gamma_{m1} & \cdots & {}_t\gamma_{mm} \end{pmatrix}. \tag{C.19}$$

If we consider the same structure in the transformation as for (C.17), then the seasonality should be modelled in all the off-diagonal elements. Assuming p parameters are needed in each of these elements, we have $m(m-1)p$ parameters to estimate. This will increase parabolic for increasing m. To limit this increase the seasonality has been modelled as follows.

Considering a model based on a m-state Markov chain $\{C_t\}$ with transition probability given by

$$Pr(C_t = j|C_{t-1} = i) = {}_t\gamma_{ij}, \tag{C.20}$$

for $i = j$, we define the working parameter

$${}_t\tau_{ii} = \boldsymbol{\beta}_i \boldsymbol{y}_t', \tag{C.21}$$

this is the part where the seasonality is incorporated. $\boldsymbol{y}_t$ is a vector of p covariates modelling the seasonality, and $\boldsymbol{\beta}_i$ is a vector of p parameters. For $i \neq j$, ${}_t\tau_{ij} \in \mathbb{R}$. Let $g : \mathbb{R} \rightarrow \mathbb{R}^+$ be a strictly increasing function such as $g(x) = \exp(x)$. Then define

$${}_t\varrho_{ij} = g({}_t\tau_{ij}), \tag{C.22}$$

and set ${}_t\gamma_{ij} = {}_t\varrho_{ij} / \sum_{k=1}^{m} {}_t\varrho_{ik}$, which is compliant with the row-sum constraint $\sum_j \Gamma_{ij} = 1$.

The transition probability matrix, for transitions between time $t-1$ and t is then given by

$$ {}_t\mathbf{\Gamma} = \begin{pmatrix} \frac{\exp(\boldsymbol{\beta}_{11} \cdot \boldsymbol{y}'_t)}{\exp(\boldsymbol{\beta}_{11} \cdot \boldsymbol{y}'_t) + \sum_{j \neq 1} \exp(\tau_{1j})} & \cdots & \frac{\exp(\tau_{1m})}{\exp(\boldsymbol{\beta}_{11} \cdot \boldsymbol{y}'_t) + \sum_{j \neq 1} \exp(\tau_{1j})} \\ \vdots & \ddots & \vdots \\ \frac{\exp(\tau_{m1})}{\exp(\boldsymbol{\beta}_{mm} \cdot \boldsymbol{y}'_t) + \sum_{j \neq m} \exp(\tau_{mj})} & \cdots & \frac{\exp(\boldsymbol{\beta}_{mm} \cdot \boldsymbol{y}'_t)}{\exp(\boldsymbol{\beta}_{mm} \cdot \boldsymbol{y}'_t) + \sum_{j \neq m} \exp(\tau_{mj})} \end{pmatrix}. \quad \text{(C.23)} $$

With this approach of modelling the seasonality in ${}_t\Gamma$, we have $m(m-1) + mp$ parameters to estimate, this is still a parabolic growth, but for $m > 2$ and $p > 1$ it is smaller than $m(m-1)p$.

An example for a model incorporating a r period seasonality is shown below:

$$ {}_t\tau_{ii} = \beta_{i1} \cos(2\pi t/r) + \beta_{i2} \sin(2\pi t/r) \text{ for } i \in \{1, ...m\}, \quad \text{(C.24)} $$

and then calculating ${}_t\Gamma$ for $t \in \{1, 2, ..., r\}$ with (C.23). Additional sine-cosine pairs can be included to model more complex seasonal patterns, if necessary. Using sine-cosine pairs, as in (C.24) is equivalent to a Fourier series without the intercept. An important note for ${}_t\tau_{ii}$ is that no intercept should be included in these, since the intercept would be confounded with the parameters ${}_t\tau_{ij}$ and thus be non-identifiable [17].

C.2.2.3 Covariates in the state-dependent distribution

For a HMM where the state-dependent distributions are gamma distributions, the conditional mean is ${}_t\mu_i = {}_tk_i\theta_i = E(X_t|C_t = i)$ where k is the shape parameter and θ is the scale parameter. This can be dependent on the vector $\boldsymbol{y}_t$ of q covariates and, for instance, described as follows:

$$ \log_t\mu_i = \boldsymbol{\alpha_i}\boldsymbol{y}'_t, \quad \text{(C.25)} $$

and then ${}_tk_i = \exp({}_t\mu_i)/\theta_i$.

It could also be considered to let covariates enter only one or some of the state-dependent distributions.

C.2.3 Model selection

C.2.3.1 Information criteria

To choose an appropriate number of states or to choose between competing state-dependent distributions, the Akaike (AIC) and the Bayesian information criteria (BIC) are used [17]. These are measures of the relative quality of a statistical model for a given set of data. A lower AIC or BIC value indicates better quality of a model relative to a given data set.

C.2.3.2 Continuous Rank Probability Score

The Continuous Rank Probability Score (CRPS) compares a probability distribution function (pdf) with an observation, where both are represented as the cumulative distribution functions (cdfs) (Figure C.2). For model S the CRPS is given as

$$CRPS(S) = \frac{1}{N} \sum_{i=1}^{n} \int_{x=-\infty}^{x=\infty} \left(F_i^S(x) - F_i^0(x)\right)^2 dx, \tag{C.26}$$

where N is the number of observations, $F_i^S(x)$ is the forecast of the cdf at the i'th observation. $F_i^0(x) = 1(x \geq x_i)$ is the indicator function for x greater or equal to the i'th observation which represents the observed cdf [21]. Observations close to the mean/steepest point of the forecast cdfs will get a low CRPS value. The CRPS is used to compare the forecasting performance between models, like comparing a homogeneous HMM with an inhomogeneous HMM. A lower CRPS value indicates better forecasting performance.

C.2.3.3 Pseudo-residuals

Beside the use of information criteria to find a suitable model, we need to assess whether the model is adequate or to assess the goodness of fit, and to detect outliers relative to the model. One way to do this is to use the pseudo-residuals [17].

For HMMs, two kinds of pseudo-residuals are useful. Those that are based on the conditional distribution given all other observations ($Pr(X_t = x_t|\mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})$), denoted ordinary pseudo-residuals [17], and those given all preceding observations ($Pr(X_t = x_t|\mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)})$); denoted forecast pseudo-residuals. Here we will consider the ordinary pseudo-residuals.

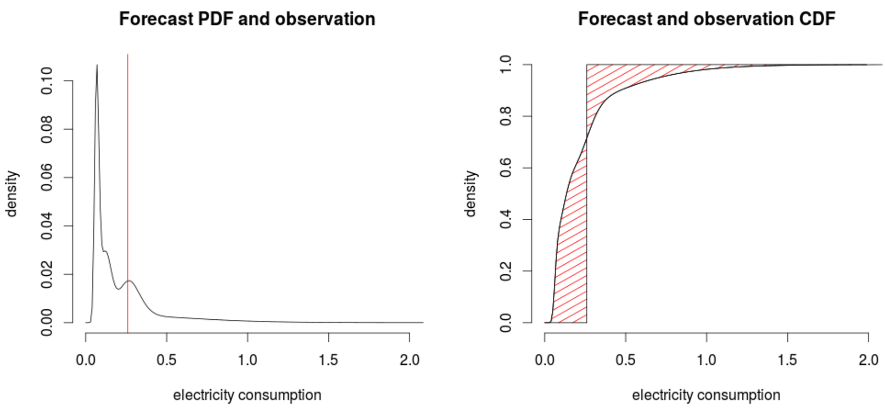


Figure C.2: Example of the cdfs used to calculate the CRPS together with pdf and observation.

For continuous observations, the ordinary pseudo-residuals are defined as

$$z_t = \Phi^{-1}\left(Pr(X_t \leq x_t | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})\right). \tag{C.27}$$

For the discrete observations of electricity consumption, we have defined the normal pseudo segment as $[z_t^-; z_t^+]$, where

$$z_t^- = \Phi^{-1}\left(Pr(X_t \leq x_t - \Delta | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})\right), \tag{C.28}$$

and

$$z_t^+ = \Phi^{-1}\left(Pr(X_t < x_t + \Delta | \mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})\right) \tag{C.29}$$

for suitable Δ.

The ordinary pseudo-residuals are used for validation of the model and for outlier detection. If the model is valid, the ordinary pseudo-residuals should be standard normal distributed. It should be noted that the pseudo-residuals might not be independent.

C.3 Results

C.3.1 Data overview

The data are collected in and nearby an apartment building in Catalonia, Spain (Figures C.19). The data consist of hourly weather data from a nearby weather station and various smart metering data from 44 apartments. The data were measured from mid July 2012 till the end of December 2013. Furthermore, occupant surveys for most of the apartments were available.

C.3.1.1 Weather data

To investigate whether the weather data (Table C.1) could be used as possible covariates/explanatory variables, an exploratory analysis was conducted on these time series. Diurnal seasonality was identified for T_a, G, W_s and W_d by using cumulative periodograms [22]. For T_a annual seasonality was also observed. This implies a possible use for these time series, namely as covariates/explanatory variables.

Table C.1: Description of weather data.

Variable	**description**
T_a	Ambient temperature in $°C$
G	Solar radiation in W/m^2
W_s	Average wind speed in m/s
W_d	Average wind direction in $°$
P	Precipitation in mm

C.3.1.2 Smart metering data

Due to a large number of zero observations, the smart metering data (Table C.2) is aggregated from 10-minute intervals to hourly intervals by summing over each hour. The space heating, hot water and water measurements are integer, since the consumption is measured in ticks. The electricity measurements are discrete with increments of 0.01. Some periods where data collection failed have been filled with the average value of these periods. These periods have been removed from the time series due to the lack of variation.

Due to the many zero observations (Table C.3) for space heating, hot water and water consumption in the aggregated data, electricity consumption is chosen as

Table C.2: Description of metering data.

Variable	**description**
x_e	Electricity consumption in kWh
x_{sh}	Space heating in kWh
x_{hw}	Hot water consumption in kWh
x_w	Water consumption in litters

the response variable for the HMMs.

Table C.3: Count of zero observation from the smart metering data.

Apartment	**2**	**5**	**18**
Electricity	0	0	20
Space heating	10433	9370	8930
Hot water	12564	11600	11643
Water	10296	7786	7363
Total number of observations	12816	12816	12816

To investigate whether there are differences between weekdays and weekends, weekly averages were calculated for one year of data (Figure C.17). For Apartment 2, no difference was observed. For Apartment 18, the midday spike is gone/reduced in the weekends. To limit the model complexity, weekly variation will not be considered in the following.

Looking at histograms (e.g. Figure C.18) of the electricity consumption, they resemble multi-modal distributions which fit the framework of the HMMs. It is suggested to use Poisson, normal, log-normal or gamma distributions in the state-dependent distributions. Poisson is suggested because the electricity consumption is discrete and could be modelled as a counting process. Using the discrete likelihood, the normal, log-normal and gamma distribution can be used.

Because the support of the normal distribution is $\mathbb{R}$, this distribution might fail in forecasting, as negative consumption might be produced (and predictive distributions will certainly include negative values). Since some of the observations are zero for the electricity consumption in some of the apartments, a log transform of these data is also not suitable.

Since the data are discrete with increments of 0.01, we can consider an observation as the interval $[x_i - 0.005, x_i + 0.005)$, when calculating the likelihood. For the zero observations, we define the intervals as $(0, 0.005)$. Then, by using the discrete likelihood on these intervals, the gamma distribution is supporting all observations.

Table C.4: The log-likelihood, AIC and BIC for model fits using Poisson and gamma distributions, Apartment 2.

Distribution	no. states	no. parameters	log-likelihood	AIC	BIC
gamma	2	6	35797	71606	71648
	3	12	31492	63007	63092
	4	20	31244	62527	62668
	5	30	31011	62081	62292
	6	42	30736	61556	61852
	7	56	30588	61288	61682
Poisson	2	4	55662	111332	111360
	3	9	37154	74326	74389
	4	16	33275	66581	66694
	5	25	31961	63971	64147
	6	36	31368	62809	63062

C.3.2 Homogeneous Hidden Markov Models

The models used for the electricity consumption are first-order Hidden Markov Models (equations C.5 and C.6), and the structure of this model is illustrated in Figure C.1.

In the previous section, we found that Poisson and the gamma distributions might be suitable. To choose the most suitable distribution, the AIC and BIC values are compared (Table C.4). From these the gamma distribution is clearly preferred. Hence, the suggested distribution for the state-dependent distribution is the gamma distribution.

The models are fitted using one year of data. Estimated parameters are the shape k_i, the scale θ_i, and the transition probability matrix Γ for $i = \{1, ..., m\}$, where m is the number of states. Only $m(m-1)$ parameters must be estimated in Γ due to the row sum constraints. The number of parameters in the HMM with gamma distributions as state-dependent distributions is $m(m+1)$. For the gamma distribution, the mean and variance are given by:

$$E[X] = k\theta \tag{C.30}$$

$$Var[X] = k\theta^2 \tag{C.31}$$

Table C.5: Estimated parameters for the three state HMM and calculated stationary distribution, mean and variance, Apartment 2.

State	k	θ	γ_{i1}	γ_{i2}	γ_{i3}	δ	Mean	Variance
1	7.74	0.012	0.85	0.14	0.01	0.52	0.09	0.001
2	7.30	0.040	0.21	0.73	0.06	0.38	0.29	0.012
3	5.14	0.205	0.00	0.30	0.70	0.10	1.05	0.216

These are used to interpret and classify the states.

C.3.2.1 Homogeneous HMM for Apartment 2

A series of models with 2-7 states was fitted to Apartment 2. The AIC and BIC values (Table C.4) point towards using many states to describe the variation in the data. Since the goal is to classify the states with respect to occupant behaviour and to limit the model complexity, due to further model development of inhomogeneous HMMs, a huge amount of states would be too difficult to interpret, and would increase the number of parameters and thus the complexity of the model. This means a suitable amount of states should be chosen. The decrease in log-likelihood, AIC and BIC, is less drastic after three states (Table C.4), hence the gain of using a four-state HMM model instead of a three-state HMM is much less than going from two to three states, hence the three-state model is investigated further.

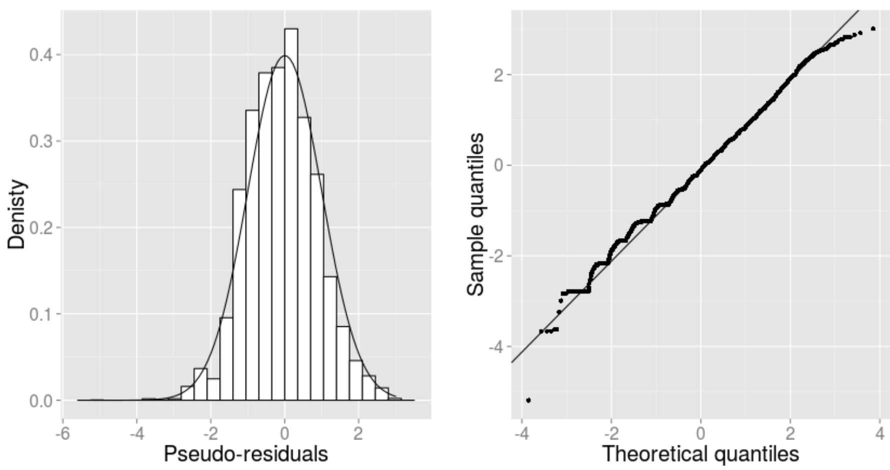


Figure C.3: Histogram and qq-plot of the ordinary pseudo-residuals for the three-state HMM, Apartment 2.

To assess the goodness of fit of the three-state model, we look at the ordinary pseudo-residuals (Figure C.3) to check if there is extreme deviance for the observation. From the plot of the histogram there does not seem to be any strong indication against the model. This is also concluded based on the qq-plot, beside that the discreteness of the data is observed. Hence, the model chosen is the three-state HMM. The parameter estimates are shown in Table C.5.

From the parameter estimates, we can interpret the states by the mean values of the state-dependent distributions. A first interpretation of the states could be:

1. Low consumption
2. Medium consumption
3. High consumption

for states 1, 2, and 3, respectively.

C.3.2.2 Global decoding

The states for the fitted models have been colour-coded in accordance with their mean values; green is low, yellow is medium and red is high consumption. This colour code is used in the rest of the paper. The most likely sequence of states is estimated using global decoding (C.9).

From the average probability profile with respect to time of day, a clear variation/dependence is observed (Figure C.4). From the average profile for time of year, there is a clear increase in the probability for state 3 in the summer months July and August. For the rest of the year, no significant differences are observed between the months. For the average probability profile for ambient temperature, the increase in temperature seems to have an increasing effect on the probability of state 3. Profiles for average wind velocity, average wind direction, precipitation and solar radiation were also produced, but no significant variation in the probability profiles was observed.

Given the average probability profile for time of day and the knowledge from the occupant survey (Table C.7), the ”low consumption” state, could be interpreted as “absent or asleep” due to the high probability for this state at night and in the evening. The “medium consumption” state could be interpreted as ”home medium consumption” due to the fact that the occupant is retired and that the probability for this state is high during the day and late evening. The ”high consumption” state could be interpreted as ”home, high consumption”, based on the knowledge that this apartment has air-conditioning and the observed temperature dependence. To summarise, the new interpretations of the states are:

1. absent or asleep

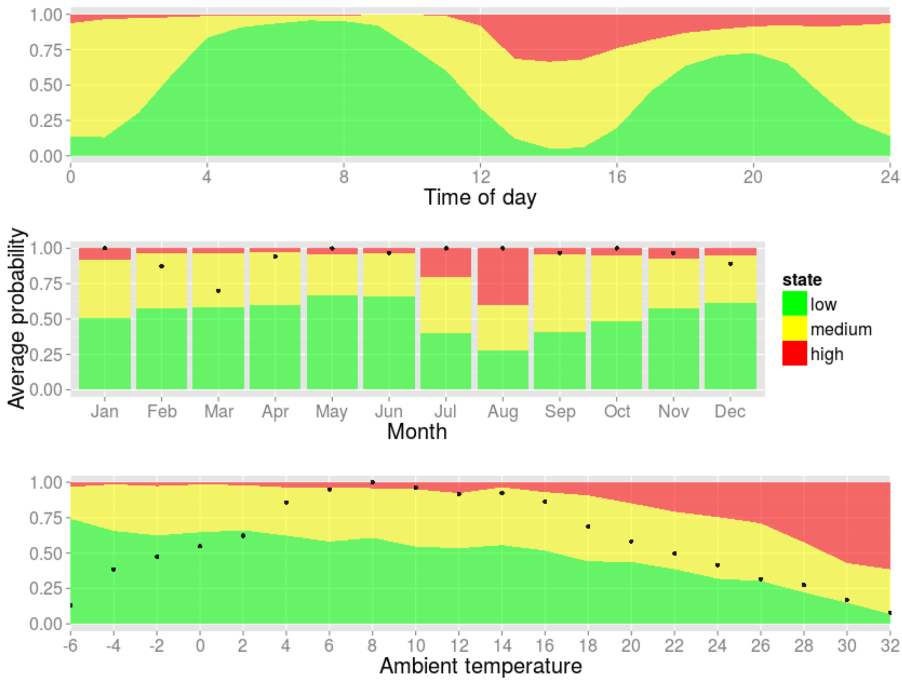


Figure C.4: Average probability profiles of being in given state, dependent on time of day, time of year and ambient temperature, Apartment 2. The black dots denote the relative amount of observations for each month or temperature interval.

2. home, medium consumption
3. home, high consumption

C.3.2.3 Comparing obtained states with other metering data

To validate the new interpretation of the states, the global decoding is compared with the data of water and hot water use. The assumption is that no water is used when the occupant is absent, hence a count of water use in the three states is done (Figure C.5), i.e. if the water-use data are greater than zero at a given time, then this is counted as water use for the given state from the global decoding at this time.

Except for the transition between when the occupant is sleeping and awake in

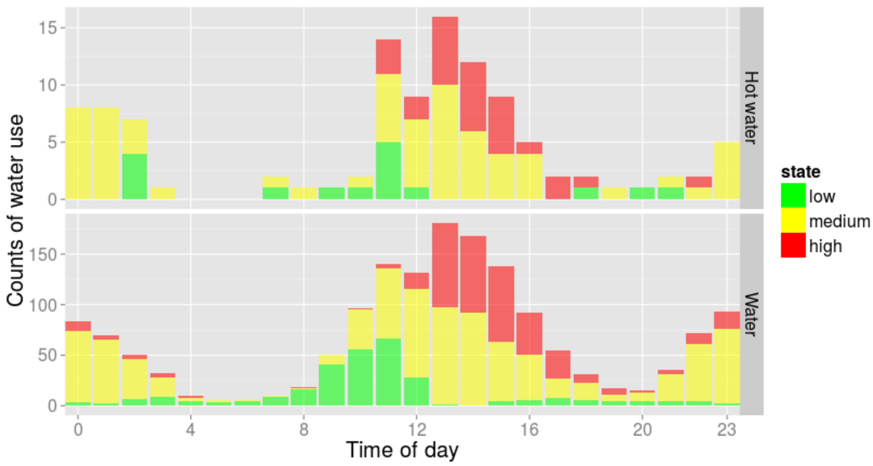


Figure C.5: Counts of water and hot-water use given state and time of day, Apartment 2.

the morning hours of the day, the probability of water use is very small for the "absent or asleep" state, hence there is no strong evidence against this interpretation of the state. For the two other states, it seems that the probability is high for water use given these states, hence the interpretation of high probability of the occupant being home given these states seems fair.

C.3.2.4 Summary of all homogeneous model fits

Data from 44 apartments was available. Many of these did not have a consistent consumption throughout the whole period of data collection and only those that did have a consistent consumption were chosen.

In total, models for 14 apartments were fitted with two-six states. For 10 of them, three state models were found suitable and for the rest, four state models were found suitable. A selection of survey data and the number of states chosen for the models are shown in Table C.7. The choice of model for each apartment is based on the AIC/BIC values and plots of the state-dependent distributions (Figure C.18), as shown for Apartment 2.

A few more apartments were fitted, but for these the optimisation of the likelihood only found local maxima for the parameters, and these results are therefore not presented.

C.3.2.5 Dependencies on explanatory variables

For all the apartments, clear diurnal dependencies were observed from the average probability profiles with respect to time of day. For some of the apartments, a dependency on ambient temperature was observed (Table C.6).

From the average probability profiles with respect to time of year, the dependency on temperature seems to originate from the summer period, mainly July and August. For Apartment 2, the knowledge of an air-conditioner seems to help explain this dependency in the high-consumption state. For the rest of the apartments with temperature dependence, the medium or low consumption states are affected, but we do not have any prior knowledge that could explain these dependencies. No significant dependences on average wind velocity, average wind direction, precipitation and solar radiation were found for any of the apartments.

Table C.6: Observation of temperature dependence (+) from the average probability profiles of ambient temperature.

State \ **Apartment**	**2**	**3**	**7**	**12**	**25**	**35**
1	-	-	-	-	-	-
2	-	+	+	+	+	+
3	+	-	-	-	-	+
4						-

Table C.7: Selection of data from occupant survey and the number of states found suitable for each apartment.

Apartment	No. occupants	Air-condition	Hours absent on weekday	Source of income	No. states
2	1	yes	$3-5$	pension	3
1	NA	NA	NA	NA	3
5	2	no	> 10	work	4
7	2	no	$3-5$	scholarship	3
26	1	no	$3-5$	pension	4
35	1	no	$6-8$	pension	4
3	3	no	< 2	pension	3
18	5	no	< 2	work	3
29	3	no	< 2	work	3
44	1	no	< 2	pension	4
12	1	no	$3-5$	pension	3
15	1	no	> 10	work	3
25	1	no	> 10	work	3
30	1	no	$6-8$	subsidy	3

C.3.2.6 Comparing average probability profiles with occupant survey

From the average probability profiles with respect to time of day, four distinct patterns were observed (Figure C.6). The four patterns have been classified as shown in Table C.8.

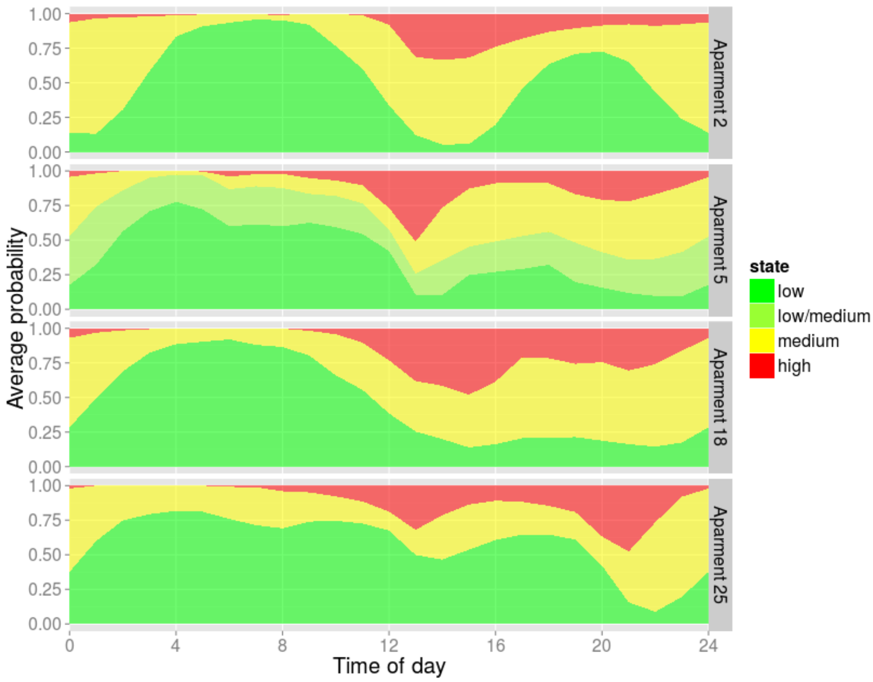


Figure C.6: Observed distinct profiles. The low/medium state of Apartment 5 is considered low based on the mean value of the state-dependent distribution.

For the four apartments with the pattern "mostly at home", the prior knowledge from the occupant survey (Table C.7), hours absent on a weekday (less than two hours for all four), corresponds well to these profiles, since they have low probability of being absent during the day. Furthermore, there might be a correspondence between low probability of being absent during the day and the fact that three or more occupants are living in the apartment, since this is the case for Apartments 3, 18 and 29.

For the "mostly absent" pattern, the prior knowledge from the occupant survey of Apartments 15 and 25 correspond well, since hours absent on a weekday are more than 10 hours. There might be a correspondence between high probability of being absent during the day and the source of income being work, since this is the case for Apartments 15 and 25. For Apartments 12 and 30, this is not the case. Common for all four apartments is that they have only one occupant.

Table C.8: Apartments classified based on the average probability profile given time of day.

Class	Apartments
afternoon/evening absence	2
equal probability for being home or absent	1, 5, 7, 26 and 35
mostly at home	3, 18, 29 and 44
mostly absent	12, 15, 25 and 30

For the pattern "equal probability for being home or absent", there is no distinct knowledge from the survey that is common for a majority of these apartments. Comparing Apartments 26 and 35, they both have one occupant who receives a pension, but the hours of absence on a weekday differ slightly. For Apartment 26 it is 3-5 hours and for Apartment 35 it is 6-8 hours. When the profiles are compared, the probability of being home during the day is slightly higher for Apartment 26 than for Apartment 35. Apartments 5 and 7 both have two occupants. The main difference from the survey is that Apartment 5 is empty for more than 10 hours a day and Apartment 7 is empty 3-5 hours a day. This is also observed when comparing the probability of being home during the day. The reason why Apartment 5 does not look like the profiles from Apartments 15 and 25 might be due to the number of occupants.

The pattern observed for Apartment 2 seems to stand out compared to the others. Two have a tendency for a higher probability of being absent during the afternoon/evening, these are Apartments 30 and 44. From the occupant survey, we know they all have one occupant and both Apartments 2 and 44 are receiving a pension.

With these observations, we have some indications of possible random or fixed effects that could enter in a population model [18], where several apartments are collected in one model with some common parameters and some apartment-specific parameters.

To investigate whether the homogeneous models have common parameters, a series of box-plots was produced for the three-state models (Figures C.7 and C.8).

For the parameter k, huge variation is observed for the low and medium consumption states (Figure C.7). For the high consumption state, there is much less variation, and this might indicate that this could enter a population model as a common parameter. For the parameter θ, the opposite is observed; here the high consumption state has huge variation between apartments, and the low and medium consumption states have low variation, which might indicate that the parameters for these two states could enter a population model as common parameters. Similar observations are seen for the four-state models.

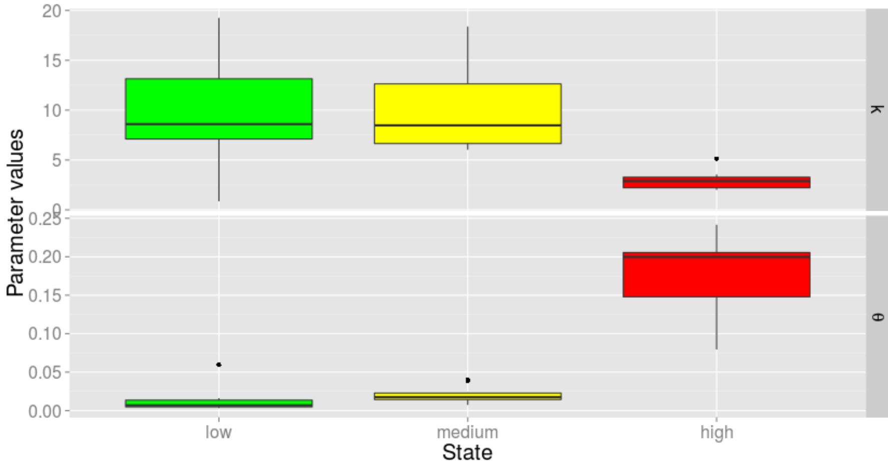


Figure C.7: Box-plots of the parameters estimates for the state-dependent distributions, note the difference in the scales.

Looking for common parameters in the transition probabilities, low variance is observed for the transition probabilities in the low consumption state, which might indicate common parameters (Figure C.8). For medium and high consumption states, a larger variance is observed, except for the transition probability of going from the high consumption state to the low consumption state which is almost zero. This indicates that the transition from high to low is highly unlikely, and in order to go from high to low, you have to go through the medium state. Similar observations were seen for the fourstate models.

C.3.3 Time-inhomogeneous Hidden Markov Models

In the time-inhomogeneous models, the time dependence is modelled in the transition probability matrix. Estimated parameters are the shape k_i, and the scale θ_i for the state-dependent distributions for $i = \{1, ..., m\}$, where m is the number of states. One year of data is used. This is the same as for the homogeneous models.

The time-dependent transition probability matrix ${}_t\Gamma$ is determined by Equation (C.23), hence the parameters are τ_{ij} for $i, j = \{1, ..., m\}$ and $i \neq j$, and $\tau_{ij} = \boldsymbol{\beta}_i \cdot \boldsymbol{y}'_t$ for $i = j$ where $\boldsymbol{\beta}_i$ is a vector of Fourier coefficients for state i and $\boldsymbol{y}_t$ is a

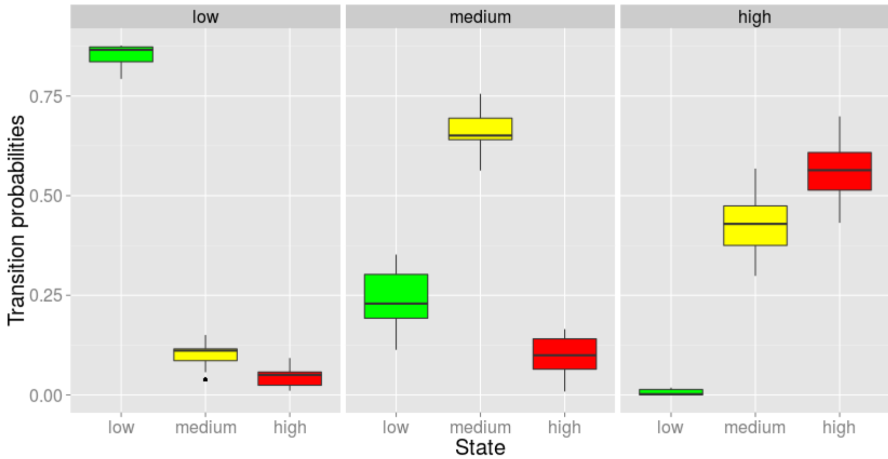


Figure C.8: Box-plots of the transition probabilities for each state.

vector of sin- cosine pairs to model the daily variation in the states. Hence the number of β parameters depends on the length of $\boldsymbol{y}_t$ and the number of states.

The Fourier series are

$$\sum_{k=1}^{p} \beta_{ik_{\sin}} \sin\left(\frac{2\pi t}{r}\right) + \beta_{ik_{\cos}} \cos\left(\frac{2\pi t}{r}\right) \quad \text{for } i = \{1, ..., m\} \tag{C.32}$$

where t is the time, r the period and p is the number of sin- cosine pairs. The model chosen is based on the AIC/BIC values and comparison of a model-generated probability profile, with the profile generated from the data in the homogeneous model fit. The number of parameters in the time-dependent inhomogeneous models is $m(m + 2p + 1)$

C.3.3.1 Time-dependent inhomogeneous HMM for Apartment 18

For Apartment 18, four models were fitted using one to four sin- cosine pairs in the Fourier series with a period of $r = 24$. The number of states found suitable for the homogeneous model fit was used, hence $m = 3$ is the number of states in these models.

From the BIC values (Table C.9) we see that two sin- cosine pairs in the Fourier series provide the best fit.

Table C.9: The log-likelihood, AIC and BIC for each fitted model of Apartment 18.

Inhomogeneous				
no. sin-cosine pairs	**no. parameters**	**log-likelihood**	**AIC**	**BIC**
1	18	31383	62801	62928
2	24	31268	62584	62753
3	30	31245	62550	62761
4	36	31237	62545	62766
Homogeneous				
	12	31710	63444	63528

Table C.10: Estimated parameters for the three-state HMM for Apartment 18

State	**k**	$\boldsymbol{\theta}$	$\boldsymbol{\beta}_{i1_{\text{sin}}}$	$\boldsymbol{\beta}_{i1_{\text{cos}}}$	$\boldsymbol{\beta}_{i2_{\text{sin}}}$	$\boldsymbol{\beta}_{i2_{\text{cos}}}$	$\boldsymbol{\tau}_{i1}$	$\boldsymbol{\tau}_{i2}$	$\boldsymbol{\tau}_{i3}$
1	6.57	0.017	0.571	2.105	-0.931	0.153	-	-2.87	-3.14
2	13.27	0.019	0.339	-0.624	0.619	-0569	-1.18	-	-1.39
3	2.85	0.21	-0.095	-5.203	2.672	0.202	-0.46	2.18	-

When checking the goodness of fit, by assessing the normal pseudo-residuals (Figure C.9) there is no strong evidence against the model from the histogram of these. From the QQ-plot, the tails are heavy and this indicates a lot of outliers for this model. This implies that the model does not capture high-consumption observations and small low-consumption observations very well. This is not a problem for classification of the states by global decoding, since e.g. a very high observation will have highest probability of being in the high-consumption state. For forecasting distributions, this is a problem since there is very low probability for high consumption using this model.

To analyse why there is a heavy lower tail, a plot of the ordinary pseudo-residual plotted as function of time is investigated (Figure C.9). From this we observe that the variation increases for the lower pseudo-residuals over time. When this is compared to the time series of the electricity consumption, we observe a simultaneous increase in the variation, implying that electricity consumption has many observations lower than before. This indicates that the occupants have changed behaviour and suggests that some kind of filtering or time-adaptive method should enter the model. The parameter estimates are shown in Table C.10.

The main difference between the homogeneous and time-inhomogeneous models is the transition probabilities. We now have the transition probabilities as a function of time of day for each state. Looking at the forecasting distributions (Figure C.10), it seems to resemble the daily probability profile when considering the states as low, medium and high consumption. The probabilities for high consumption (1 to 3 kWh) are very low, this might explain the heavy upper tail

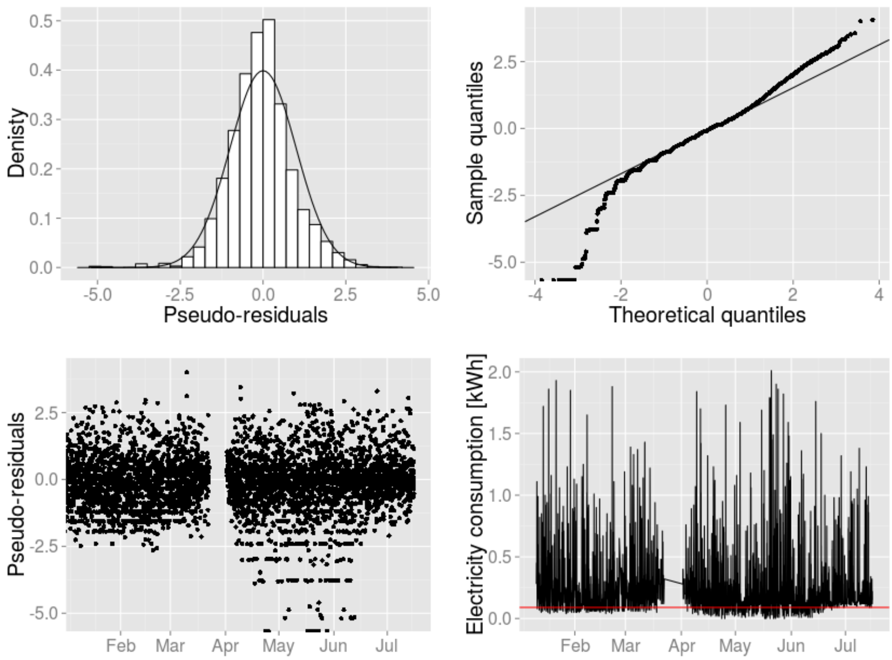


Figure C.9: Histogram and qq-plot of the ordinary pseudo-residuals for the three-state inhomogeneous HMM of Apartment 18, also plotted as a function of time. The electricity consumption of the same period of time is shown with the mean of the low consumption state.

in the QQ-plot of the ordinary pseudo-residuals, since there are observations in the electricity consumption for Apartment 18 on up to 3 kWh, which has very low probability in the forecasting distributions.

To check whether the inhomogeneous model is a better fit than the homogeneous model, we first look at the AIC and BIC values (Table C.9). It is observed that there is an improvement of the model with added time dependence.

To evaluate the forecasting performance, we use the forecasting distribution $Pr(X_{T+h} = x|\mathbf{X}^{(T)} = \mathbf{x}^{(T)})$ for $h = \{1, 2, ..., 48\}$, where $\mathbf{x}^{(T)}$ is the data used to fit the model. The CRPS is then calculated for the 48 forecast horizons using the 48 observations after $\mathbf{x}^{(T)}$. This is done for the remaining half-year of data and we then obtain around 4000 CRPS intervals with 48 horizons. The mean value for each horizon is calculated. This is done for both the homogeneous and

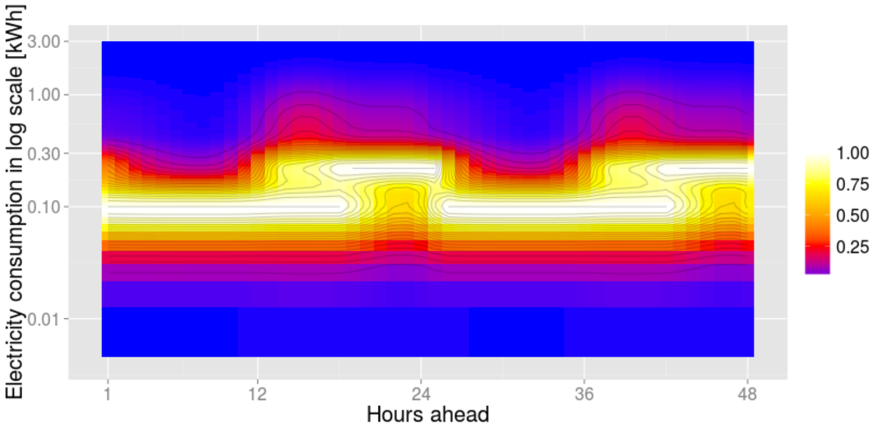


Figure C.10: Contour plot of forecasting distributions 48 hours ahead of the data used to fit the model. The scale is relative to the largest probability in each horizon, Apartment 18.

the time-inhomogeneous model and results are then compared (Figure C.11). It is observed that for horizon 1, they seem to be close, which is expected due to the nature of a first-order Markov chain. When comparing the mean values for the rest of the horizons, we observe a clear difference between the two models, where the time-inhomogeneous model has the better performance.

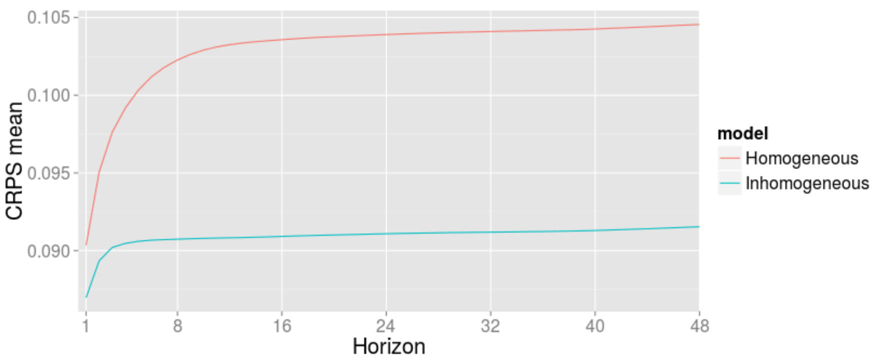


Figure C.11: Comparison of the CRPS between homogeneous and time-inhomogeneous models, Apartment 18..

Overall, we have observed that time-dependence transition probabilities lead to better forecasting performance, and hence that the inhomogeneous models are to be preferred.

C.3.3.2 Summary of all inhomogeneous model fits

For all the apartments, time-inhomogeneous models were fitted with one to four harmonic pairs in the Fourier series in the transition probabilities, as in Table C.9. For the three-state models, two harmonic pairs were found adequate, and for the four-state models three pairs were found adequate. For the models where temperature dependence was found in the homogeneous model (Table C.6), the model-generated daily probability profiles are less accurate than for the apartments were no temperature dependence was found. For Apartments 3 and 7, they do not resemble the homogeneous profile at all. This might indicate that the temperature should enter these models.

Only small decreases are seen in the AIC/BIC values for all the model fits adding extra sin- cosine pairs in the transition probability. When comparing AIC/BIC values with the homogeneous models, improvements are seen for all the inhomogeneous models. When comparing the forecast performance with the CRPS, a similar result as for Apartment 18 is observed for all the apartments, i.e. the inhomogeneous models have better forecast performance.

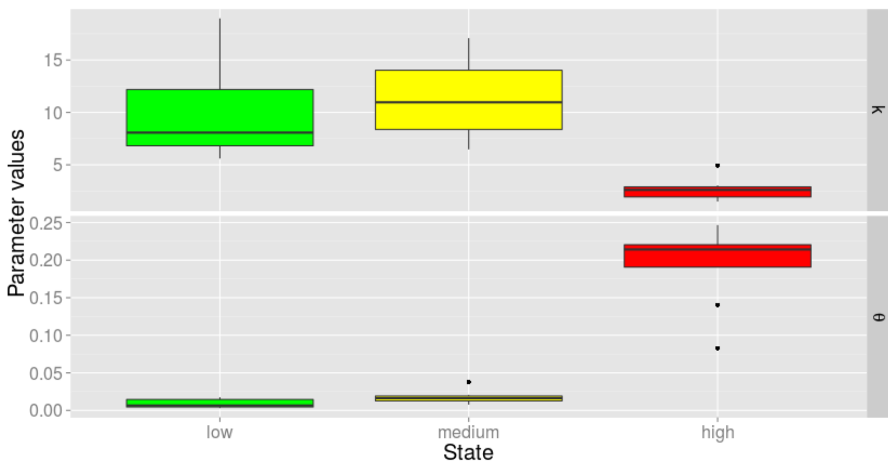


Figure C.12: Box-plots of state-dependent parameters for the inhomogeneous three-state models.

Investigating for common parameters, similar observations are seen for the state-dependent distribution parameter k and θ (Figure C.12), as for the homogeneous model fits.

For the τ parameters in the transition probabilities, the parameters for the low and medium state (denoted (1,j) and (2,j) in the box-plot) have smaller variance than the ones for the high consumption state. For the Fourier coefficients, there seems to be a lot of variation in each state and also between states. When comparing the transition probabilities for each state between apartments, very similar patterns are observed for many of the apartments. These observations might indicate that some of the parameters that determine the transition probabilities are common for many of the models.

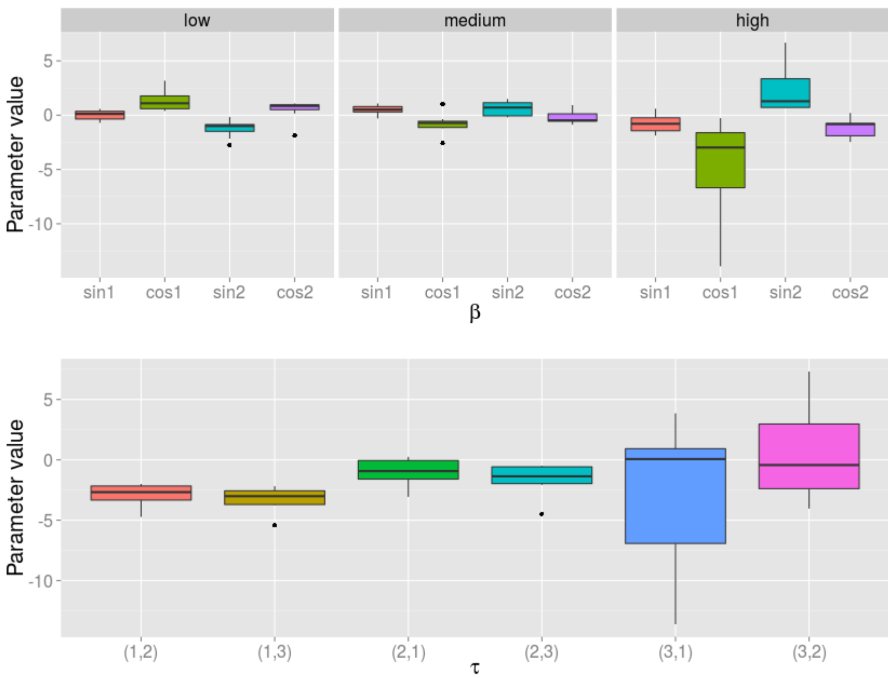


Figure C.13: Box-plots of transition probability parameters for the inhomogeneous three-state models.

In general, the forecasting distributions resemble the daily profiles well, but the very high consumption observations are not captured by most of the models, as seen for Apartment 18.

C.3.4 Time inhomogeneous temperature-dependent Hidden Markov Models

The models are basically the same as the time-inhomogeneous model. The temperature dependence is modelled in q states, then for these states the conditional mean is modelled as in (C.25), where only the temperature is used as covariate. Hence for each state dependent on temperature, the means are given by

$$\log_t \mu_i = \alpha_{i1} + \alpha_{i2} y_t \tag{C.33}$$

where i is the state dependent on temperature and y_t is the temperature at time t. Given this implementation, the number of parameters to be estimated is $m(m + 2p + 1) + q$

C.3.4.1 Time-inhomogeneous temperature-dependent HMM for Apartment 2

For Apartment 2, temperature dependence was observed in the high consumption state, hence the temperature dependence is modelled in this state.

It is observed that the mean value of this state shows an annual variation (Figure C.14) with significantly higher mean values in the summer period than the winter period.

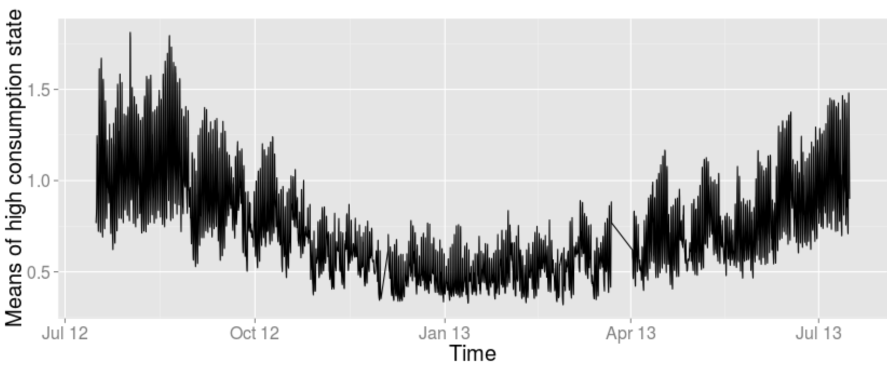


Figure C.14: Mean values of the high consumption state over one year.

To check whether the temperature-dependent model is a better fit than the homogeneous and time inhomogeneous models, we first look at the AIC and BIC values (Table C.11). It is observed that there is an improvement of the model with added time and temperature dependence, though it is a smaller improvement from the time-dependent to the temperature-dependent model than from the homogeneous to the time-dependent model.

Table C.11: Comparison of Homogeneous and Inhomogeneous models, Apartment 2.

Model	no. parameters	log-likelihood	AIC	BIC
Homogeneous	12	31492	63007	63092
Inhomogeneous	24	30653	61355	61524
Temp. dependent	25	30461	60973	61149

When the CRPS is compared (Figure C.15), we also see an improvement in the forecasting performance for the time- and temperature-dependent model.

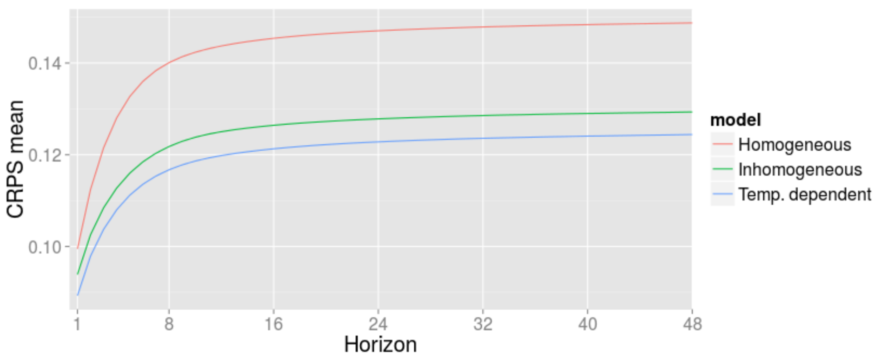


Figure C.15: CRPS comparison between the homogeneous model, time-inhomogeneous and the time-inhomogeneous temperature-dependent models, Apartment 2.

When the forecasting distributions are compared (Figure C.16), the differences between these models are observed. Clear variances from day to day are seen in the probabilities for the high consumption in the temperature-dependent model.

Temperature-dependent models where also fitted for Apartments 3, 7, 12, 25 and 35, but did not yield good results. This might be due to the huge variation in the mean values during the day, so a smoothing of these mean values is suggested. Apartment 2 is less affected by this, since the probability of being in the high consumption state is very low during most of the day, except for the hours in the middle of the day.

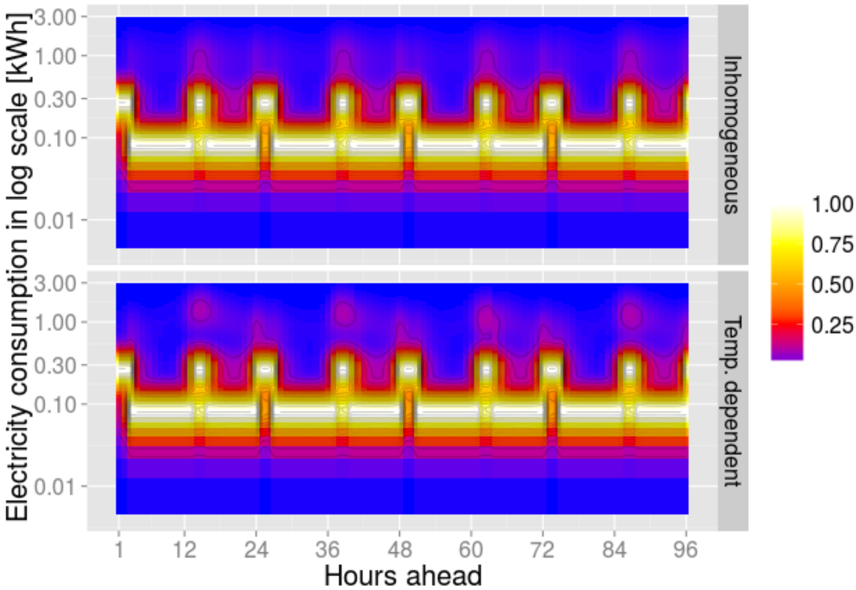


Figure C.16: Contour plot of forecasting distributions for the inhomogeneous and temperature-dependent model, Apartment 2. The scale is relative to the largest probability in each horizon. The forecast is in the summer period.

C.4 Discussion/Conclusion

By applying homogeneous Hidden Markov Models on frequent measurements of electricity consumption from smart metering data, we have classified the obtained states in accordance with occupant behaviour by global decoding for fourteen apartments. The classification is:

- ”low consumption” and ”absent or asleep”
- ”medium consumption” and ”home”
- ”high consumption” and ”home, high consumption”

Given the hourly observations of the electricity consumption, used in this study, the states have not been classified to specific household appliances. This is because a sequence of different appliances could have been used during one hour.

We have found diurnal dependencies for all the apartments and ambient-temperature dependence for several. Hence, the applications of the homogeneous models are to classify the states in accordance with occupant behaviour and to identify possible covariates/explanatory variables. With the sequence of the global decoding, it is possible to select data from other smart metering data, dependent on a certain state. As an example, space heating data could be investigated for conditions when the most probable state is "absent or asleep". By examining the average probability profiles with respect to time of day for the apartments, four distinct patterns of daily occupant behaviour were observed. This suggests that we might be able to classify the daily occupant patterns of the apartments.

These classifications of the states are relative to the individual apartments and the levels are determined by the underlying distribution describing the consumption in an individual state. Hence, the level of low, medium and high consumption varies between apartments. Based on the diurnal average probability profile and the levels of the states, consumers can optimize their consumption and save money. The utilities can use these models quantitatively to predict load forecasts in blocks, streets, cities or regions and use these forecasts in trading electricity or to discover peak loads more locally to optimise the dimension of the grid.

Due to the fact that there are possibilities for multiple changes of states, between successive observations (hourly), it is suggested to investigate the application of continuous-time Markov chains in the HMMs. This might also help to explain the number of residents who are present in the apartments, by adding more states in the HMMs leading to new interpretations of these. In this case the continuous time formulation might lead to a parametrization using less parameters than the corresponding discrete-time Markov chain.

For the time-inhomogeneous models, we found that in general the forecasting distributions resembled the homogeneous data-driven probability profiles well. From the comparison of the homogeneous and time-inhomogeneous models by CRPS, a clear improvement of the forecasting for the time-inhomogeneous models was observed. The application of the time-inhomogeneous models in this paper is to forecast electricity consumption and simulate occupants' states.

In the implementation of these models, the number of parameters was increasing rapidly with the number of states. This resulted in a big increase in computer execution time.

Given models with many parameters, the optimisation of the likelihood could find local maxima instead of global. This can be countered by choosing the start values of the parameters for the optimisation in a clever way. This is very cumbersome and will not always yield the desired outcome.

In the validation of the models, we observed that for some of the apartments the occupants have tendencies to change their behaviour slightly over the year. This is not captured in the models and a solution to this could be to try some adaptive method, by weighting the observations in the model fit or by implementing seasonal varying coefficients in the state-dependent distributions.

We also observed that, for all the models, the very high consumption was not captured. The main reason for this is that the state-dependent distributions for high consumption are inadequate in capturing this due to a very high variation in the observations of these.

For the time-inhomogeneous temperature-dependent model of Apartment 2, we observed a small increase in the performance due to better forecasting in the high consumption state. For the rest of the apartments, a smoothing of the temperature-dependent mean values is suggested.

For the homogeneous models, we found clear indications of common model parameters. By analysing the different occupant patterns, several variables were found that might help describing the variation in the observations using fixed or random effect terms in the models. This is a good indication of a plausible further development of the homogeneous models as population models. For the time-inhomogeneous models, we also found common parameters, which also suggests developing these further as population models.

We found some issues (e.g. computational) which must be considered and solved for the time-inhomogeneous models and the time-inhomogeneous temperature-dependent models before taking the step to population models.

Acknowledgements

We gratefully acknowledge the financial support received for this work from CITIES (DSF - 1305-00027B).

References C

[1] P. D. Andersen et al. "Dynamic modeling of presence of occupants using inhomogeneous Markov chains". In: *Energy and Buildings* 69 (2014), pp. 213–223.

[2] S. D'Oca et al. "Effect of thermostat and window opening occupant behavior models on energy use in homes". In: *Building Simulation.* Vol. 7. 6. Springer. 2014, pp. 683–694.

[3] C. Wang, D. Yan, and Y. Jiang. "A novel approach for building occupancy simulation". In: *Building simulation.* Vol. 4. 2. Springer. 2011, pp. 149–167.

[4] J. Page et al. "A generalised stochastic model for the simulation of occupant presence". In: *Energy and buildings* 40.2 (2008), pp. 83–98.

[5] K. Sun et al. "Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration". In: *Building and Environment* 79 (2014), pp. 1–12.

[6] F. Haldi and D. Robinson. "Interactions with window openings by office occupants". In: *Building and Environment* 44.12 (2009), pp. 2378–2395.

[7] F. Haldi and D. Robinson. "On the behaviour and adaptation of office occupants". In: *Building and environment* 43.12 (2008), pp. 2163–2177.

[8] H. B. Gunay et al. "On adaptive occupant-learning window blind and lighting controls". In: *Building Research & Information* 42.6 (2014), pp. 739–756.

[9] J. Tanimoto and A. Hagishima. "State transition probability for the Markov Model dealing with on/off cooling schedule in dwellings". In: *Energy and Buildings* 37.3 (2005), pp. 181–187.

[10] U. Wilke et al. "A bottom-up stochastic model to predict building occupants' time-dependent activities". In: *Building and Environment* 60 (2013), pp. 254–264.

[11] V. Fabi et al. "A methodology for modelling energy-related human behaviour: Application to window opening behaviour in residential buildings". In: *Building Simulation.* Vol. 6. 4. Springer. 2013, pp. 415–427.

[12] Z. Yu et al. "A systematic procedure to study the influence of occupant behavior on building energy consumption". In: *Energy and Buildings* 43.6 (2011), pp. 1409–1417.

[13] Z. J. Yu, F. Haghighat, and B. C. Fung. "Advances and challenges in building engineering and data mining applications for energy-efficient communities". In: *Sustainable Cities and Society* (2015).

[14] D. Aerts et al. “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 (2014), pp. 67–78.

[15] J. Zhao et al. “Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining”. In: *Energy and Buildings* 82 (2014), pp. 341–355.

[16] R. K. Andersen. “The influence of occupants’ behaviour on energy consumption investigated in 290 identical dwellings and in 35 apartments”. In: *10th International conference on healthy buildings*. 2012.

[17] W. Zucchini and I. L. MacDonald. *Hidden Markov models for time series: an introduction using R*. CRC Press, 2009.

[18] H. Madsen and P. Thyregod. *Introduction to general and generalized linear models*. CRC Press, 2011.

[19] J. Bloem et al. “Dynamic Methodology for the Evaluation of Occupant Behaviour and Residential Energy Consumption”. In: *Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting*. 2015, pp. 591–599.

[20] M. Pinsky and S. Karlin. *An introduction to stochastic modeling*. Academic press, 2010.

[21] H. Hersbach. “Decomposition of the continuous ranked probability score for ensemble prediction systems”. In: *Weather and Forecasting* 15.5 (2000), pp. 559–570.

[22] H. Madsen. *Time series analysis*. CRC Press, 2007.

C.5 Appendix

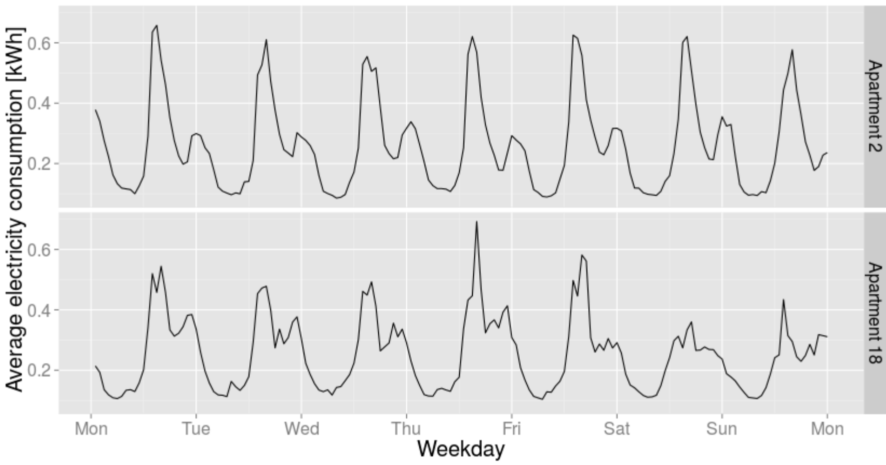


Figure C.17: Weekly averages of the electricity consumption, one year of data was used.

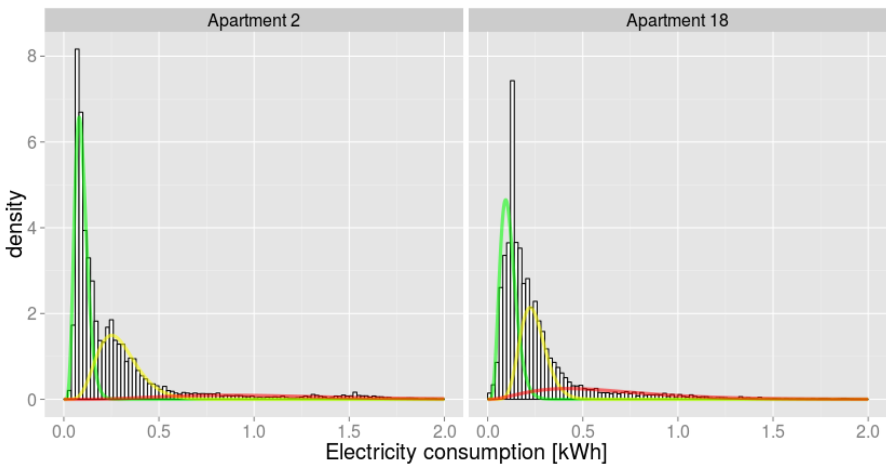


Figure C.18: Histograms of the electricity consumption, one year of data was used. Observations between 2-3 kWh are not shown.

Figure C.19: The Apartment Building

Paper D

Disaggregation using Continuous Time Hidden Markov Models applied on hourly electricity consumption

Authors:
Jon Liisberg, Jan Møller, Peder Bacher, Joakim Munkhammar, Joakim Widén, Henrik Madsen

Submitted to:
Applied Energy.

Disaggregation using Continuous Time Hidden Markov Models applied on hourly electricity consumption

Jon Liisberg[12], Jan Møller[1], Peder Bacher[1], Joakim Munkhammar[3], Joakim Widén[3], Henrik Madsen[1]

Abstract

With the increasing use of smart meters there is a rising potential to make residential electricity consumers more aware of how they are using electricity. To facilitate this, models that can disaggregate hourly electricity consumption are needed. In this paper, we present a novel approach for disaggregation of residential electricity consumption using Continuous Time Hidden Markov Model (CT-HMM). This approach is tested on a time-series of hourly observed electricity consumption of a single household. By determining the most likely sequence of states conditioned on the observations (global decoding), we observed a distribution of proportions that provides information about the probability of being in a given state at a given time of day. With this information and time-series of consumption on appliance levels, we have described the states in accordance to groups of activities in the household related to appliance use. The descriptions of the states were compared to a time-use diary, which showed a good correspondence between the state descriptions and the reported activities.

D.1 Introduction

Disaggregation is the extraction of appliance-level data from an aggregate or whole building/household energy signal using statistical approaches [1]. This is also known as non-intrusive load monitoring (NILM) [2]. Disaggregation is

[1]DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark
[2]SEAS-NVE, Digital Innovation, Hovedgaden 36, DK-4520 Svinninge, Denmark
[3]Built Environment Energy Systems Group (BEESG), Solid State Physics, Department of Engineering Sciences, Uppsala University, P.O. Box 534, Uppsala SE-751 21, Sweden

different from decomposition of a time-series where the signal is usually decomposed into trend, seasonality and remainder [3]. Usually disaggregation algorithms split up an aggregated load signal by detecting appliance signatures, i.e., patterns in the signal that are unique to the appliance. Algorithms are either supervised where labelled data of appliances/appliance signatures are needed for training, or unsupervised where no labelled data or a priori information is needed [2]. These approaches can be referred to as fine grained disaggregation [4], since data with high-resolution in the temporal domain is needed (≤ 1 minute). Coarse grained disaggregation is when lower temporal resolutions of power or power consumption in time steps from 15 minutes to 1 hour is used, usually disaggregating into loads that correlate with outdoor temperature, continuous loads and time dependent loads [1, 4].

With the implementation of the supplier-centric DataHub in Denmark [5], the availability of smart metering data of electricity consumption has increased. With this data, there are new potentials of increasing awareness among residential electricity consumers about their consumption, but the most common resolution available is hourly data on total consumption, mainly for billing purposes. For the newly installed smart-meters in Denmark, the minimum requirement for temporal resolution is 15 minutes but utilities are only required to upload hourly measurements to the DataHub [6]. In EU the smart meter roll out is unfolding and several other EU countries are implementing similar datahubs as in Denmark [7, 8], but there is no common agreement on the minimum temporal resolution only a recommendation of a minimum of 15 minutes resolution [9]. Hence, without additional equipment to the smart-meters it is unlikely that data with high resolution in the temporal domain will be available in the DataHubs.

Research on disaggregation of electricity consumption began over thirty years ago [4, 10, 11], and within the last decade a dramatic increase in number of papers on energy disaggregation have been published, see the reviews [2, 4]. In [1] several studies are reviewed, and it is found that providing disaggregated feedback of electricity consumption at appliance level leads to savings on the electricity bill, both through renovation, settings and changes in use patterns. [1] also reviews studies of disaggregation approaches. In these studies, a wide range in different current, voltage, power level resolutions and frequencies were considered. Only one of the thirty-two studies reviewed, [12], was coarse grained disaggregation using hourly measurements of power consumption. This shows that there has been more focus on fine grained disaggregation in the past but coarse grained disaggregation is slowly gaining interest, e.g., [13].

Looking into studies on disaggregation of hourly smart-meter data, we found four studies [12, 13, 14, 15]. In [12] via discriminative sparse coding, models for each appliance are trained and then used to separate an aggregated signal. This approach is supervised requiring data on appliance level for training. In [13] supervised K-Nearest Neighbours is used to disaggregate both in 15 minutes and hourly time resolution. In [14] whole-house electricity use was disaggregated, using the external temperature, into base load, activity load, cooling and heating season gradient. In [15] regression models, with weather data and response data from a household survey as input, are used to disaggregate the consumption into a modelled heat load and other appliances. Only [14] did not use prior knowledge or appliance load time series and found 3 categories of loads, base load, activity load and temperature load.

Given that a vast body of data of total household consumption without prior knowledge on appliances or time-series on individual appliances is building up in the DataHub, we believe that more research on unsupervised coarse grained disaggregation of activities and base load is needed, particularly on data with hourly resolution.

In this paper we address coarse grained disaggregation of the electricity consumption without temperature influence, i.e., the activity and base load. Hence, we propose a top-down approach for disaggregation of electricity consumption only using hourly observations, since knowledge on appliances and time-series of appliance level is not available from the DataHub. This builds on work from [16], where a simple disaggregation of electricity consumption was shown in accordance to occupant behaviour using discrete time HMM and [17, 18], where Continuous-Time Hidden Markov models (CT-HMMs) were utilised on time-series observed in discrete-time. It is argued that the CT approach significantly reduced the number of parameters needed.

The focus of the study presented in this paper is to investigate the application of CT-HMMs on hourly observations of electricity consumption in residences. The study seeks to test the hypothesis that, by applying CT-HMM on hourly observations of electricity consumption, we can split the consumption into several states, that can be interpreted as appliance related activities in the household. The CT approach also has the advantage of handling missing data or irregular observed data.

The aim is to present a modelling framework for CT-HMMs on hourly observations of electricity consumption, and then apply this framework in a single residential household test case. We focus on interpreting the states of the CT-HMMs in accordance with appliances and/or activities related to appliance use, based on the underlying appliance loads. To validate the state descriptions we compare with a time use diary of the household, and suggest further development of these models.

The outline of this paper is as follows. In Section D.2, the methods used for model fitting, validation and the key inference tool, global decoding, are described. Section D.3 describes the data of the test case household. In Section D.4 the results and validation of the model fit are outlined, then the state descriptions and the detailed reasoning for these descriptions are explained, followed by a comparison of the state descriptions with a time use diary of the test case household. Finally, in Section D.5, the results, further model development and applications of the model framework are discussed.

It is recommended to read this paper in colour for optimal readability.

Nomenclature

AIC	Akaike information criterion
BIC	Bayesian information criterion
HMM	Hidden Markov Model
CT-HMM	Continuous-Time Hidden Markov Model
cdf	cumulative distribution function
pdf, $p(x)$	probability mass or density function
m	number of states
t, s	a time stamp in discrete or continuous time
T	maximum of t, i.e., $t \in \{1, \dots, T\}$ or $t \in [0, T]$
$\mathbb{N}$	the natural numbers
$\mathbb{R}$	the real numbers
$i, j, k \in \mathbb{Z}$	integers
C_t	the state of a Markov chain at time t
X_t	the state of the random process $\{X_t\}$ at time t
x_t	the observation of the random process $\{X_t\}$ at time t
$\boldsymbol{A}, \boldsymbol{\Gamma}$	matrices
$\boldsymbol{a}, \boldsymbol{\theta}$	vectors

D.2 Methods

This section contains a brief introduction to Hidden Markov Models (HMM) based on [16, 19] and Continuous Time Hidden Markov Models (CT-HMM) based on [17, 18, 20]. A description of the methods used in the implementation and validation of the CT-HMM is also presented.

D.2.1 Discrete Hidden Markov Models

Hidden Markov Models consist of a finite number of components, and a mixing distribution. For m components, the mixture distribution depends on a component distribution, $\{\delta_1, \ldots, \delta_m\}$, and m probability or density functions, $p_1(x), \ldots, p_m(x)$ [19].

The components are specified by the discrete random variable C which performs the mixing, where $Pr(C = i) = \delta_i$ for $i \in \{1, \ldots, m\}$ and $\sum_{i=1}^{m} \delta_i = 1$. Let X denote the random variable with the mixture distribution. Then the probability or density function of X is given by

$$p(x) = \sum_{i=1}^{m} \delta_i p_i(x). \tag{D.1}$$

A sequence of discrete random variables $\{C_t : t \in \mathbb{N}\}$ is a discrete-time Markov chain if, for all $t \in \mathbb{N}$, the Markov property is satisfied, i.e.,

$$Pr(C_{t+1}|C_t, \ldots, C_1) = Pr(C_{t+1}|C_t). \tag{D.2}$$

The conditional probabilities, $Pr(C_{s+k} = j|C_s = i)$ are called transition probabilities.
The Markov chain is called homogeneous if the transition probabilities are not dependent on time, otherwise it is inhomogeneous. The k-step transition probability for a homogeneous Markov chain is denoted as

$$\gamma_{ij}(k) = Pr(C_{s+k} = j|C_s = i). \tag{D.3}$$

In particular, $\gamma_{ij}(1)$ is denoted γ_{ij} and can be collected in the transition probability matrix $\mathbf{\Gamma}$

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix}. \tag{D.4}$$

It can be shown that, for a homogeneous Markov chain, $\mathbf{\Gamma}(k) = \mathbf{\Gamma}^k$. Let $\mathbf{X}^{(T)}$ denote $(X_1, \ldots, X_T)$ and $\mathbf{C}^{(T)}$ denote $(C_1, \ldots, C_T)$. Collecting both parts, a first order HMM can be summarized by

$$Pr(C_t|\mathbf{C}^{(T-1)}) = Pr(C_t|C_{t-1}), t = 2, 3, \ldots \tag{D.5}$$

$$Pr(X_t|\mathbf{X}^{(T-1)}, \mathbf{C}^{(T)}) = Pr(X_t|C_t), t \in \mathbb{N}. \tag{D.6}$$

Where (D.5) is the Markov property. Hence, the dynamics are described by the unobserved parameter process $\{C_t : t = 1, 2, \ldots\}$, which describes the evolution of the states in time. The observations are described by the state-dependent process $\{X_t : t = 1, 2, \ldots\}$ such that when C_t is known, the distribution of X_t only depends on the current state C_t.

For the transition probability matrix $\mathbf{\Gamma}$, there are m^2 entries, but only $m(m-1)$ free parameters due to the row-sum constraint. In total we end up with $m(m-1)+mp$ parameters where p is the number of parameters in the individual state-dependent distributions. This is a quadratic growth of parameters with m, and hence we want to investigate different ways to parametrize Γ.

D.2.2 Continuous Time Hidden Markov Models

For continuous-time Markov chains, transitions may occur at all times. We will denote the transition probability matrix $\boldsymbol{P}(t)$ in the continuous case and the quantities of interest become

$$p_{ij}(\Delta t) = Pr(C(t + \Delta t) = j|C(t) = i). \tag{D.7}$$

For $\Delta t \to 0$, clearly $p_{ij}(0) = 0$ for $i \neq j$ and $\lim_{t\to 0} \boldsymbol{P}(t) = \boldsymbol{I}$. Assuming $p_{ij}(\Delta t)$ is differentiable at 0, the transition rate can be defined as

$$p'_{ij}(\Delta t) = \lim_{\Delta t \to 0} \frac{p_{ij}(\Delta t) - p_{ij}(0)}{\Delta t} \tag{D.8}$$

$$= \lim_{\Delta t \to 0} \frac{Pr(C(t + \Delta t) = j|C(t) = i)}{\Delta t} = q_{ij}. \tag{D.9}$$

Let $q_{ii} = -q_i = -\sum_{j \neq i} q_{ij}$ then $\boldsymbol{Q} = \{q_{ij}\}$ has non-negative off diagonals and non-positive diagonal elements. $\boldsymbol{Q}$ is called the transition intensity matrix.

The sojourn time or holding time is the expected time a continuous-time Markov chain will stay in a given state. These are usually modelled as exponential distributions with rate parameter q_i, and hence the mean and variance of the sojourn time, are given by

$$\mathrm{E}[X] = 1/q_i \tag{D.10}$$

$$\mathrm{Var}[X] = 1/q_i^2 \tag{D.11}$$

for $i \in \{1, \ldots, m\}$.

Given $\boldsymbol{Q}$ the transition probability matrix $\boldsymbol{P}(t) = \{p_{ij}(t)\}$ can be found by solving Kolmogorov's differential equation

$$\frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{P}(t)\boldsymbol{Q}. \tag{D.12}$$

With initial condition $\boldsymbol{P}(0) = \boldsymbol{I}$ the solution is

$$\boldsymbol{P}(t) = e^{\boldsymbol{Q}t}\boldsymbol{P}(0) = e^{\boldsymbol{Q}t}. \tag{D.13}$$

When the process enters state i, it stays there according to the sojourn time, before it jumps to another state $j \neq i$ with probability q_{ij}/q_i.

Assuming that, in a short time interval, it is only possible to change to the neighbouring states, we have

$$\begin{aligned} p_{ij} &= o(\Delta t), |i-j| \geq 2, \\ p_{ii} &= 1 - q_i \Delta t + o(\Delta t), \\ p_{i,i-1} &= w_i q_i \Delta t + o(\Delta t), \\ p_{i,i+1} &= (1-w_i) q_i \Delta t + o(\Delta t), \\ i &\in \{1, \ldots, m\}, \end{aligned} \tag{D.14}$$

where $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$ and we define state $m+1 =$ state 1 and state $0 =$ state m. It should be noted that the process cannot go directly from state i to state $i+2$ without going through state $i+1$, but there is no limit to how fast a transition from state i to state $i+2$ can occur. Under this assumption, the matrix of transition intensities has the structure

$$\boldsymbol{Q} = \begin{pmatrix} -q_1 & (1-w_1)q_1 & 0 & \cdots & 0 & w_1 q_1 \\ w_2 q_2 & -q_2 & (1-w_2)q_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ (1-w_m)q_m & 0 & 0 & \cdots & w_m q_m & -q_m \end{pmatrix}. \tag{D.15}$$

It is seen that now we end up with $2m + mp$ parameters, i.e., growing linearly with m.

D.2.3 Optimization and model choice

The model fitting is carried out using the R-package[4] msm[21] based on direct numerical maximisation of the likelihood function. In this case, the nlm function is used as the optimization routine which minimises the negative log-likelihood. Due to the possibility of local minima for the optimization, the result of the optimization should be checked. If msm returns a successful model fit, there are three possible codes describing the result from the nlm function. It should be noted that the data is log-transformed for the model fitting to avoid constrained optimization of the underlying distributions, since no negative electricity consumption is possible and it is assumed that the consumption is strictly positive in a normal household unless there is a power failure. The possible codes are:

1. Relative gradient is close to zero, current iterate is probably a solution.
2. Successive iterates within tolerance, current iterate is probably a solution.
3. Last global step failed to locate a point lower than estimate. Either estimate is an approximate local minimum of the function or *steptol* is too small.

Only codes 1 and 2 are considered as viable results, since code 3 is either a local minimum or the algorithm is stuck because the step-size has decreased below the *steptol* which is a tolerance for the minimum step size in each iteration of the optimization. In this implementation the default setting is used.

To choose an appropriate number of states, the Akaike (AIC) and the Bayesian information criteria (BIC) are used [19]. These are measures of the relative quality of a statistical model for a given set of data. A lower AIC or BIC value indicates better quality of a model relative to a given data set.

D.2.4 Initial parameter selection

Due to the assumption that in a small time interval it is only possible to change to a neighbouring state in terms of the ordering in the transition intensity matrix (D.15), it is required that the states are also neighbouring in terms of their underlying distributions, i.e., the distributions of neighbouring states need to be overlapping. This requires introduction of transition states from the distribution with the lowest mean to the highest. This is not required in a discrete

[4] https://www.r-project.org/

HMM. However, this also helps to make the resulting continuous time Markov chain recurrent. Furthermore, it is important to choose plausible starting values for the state-dependent distributions. This is critical due to the high dimension solution space for the optimization and the risk of numerical overflow or underflow if a state distribution contains few or no observations. An example of a feasible initial set of distributions is shown in Figure D.1.

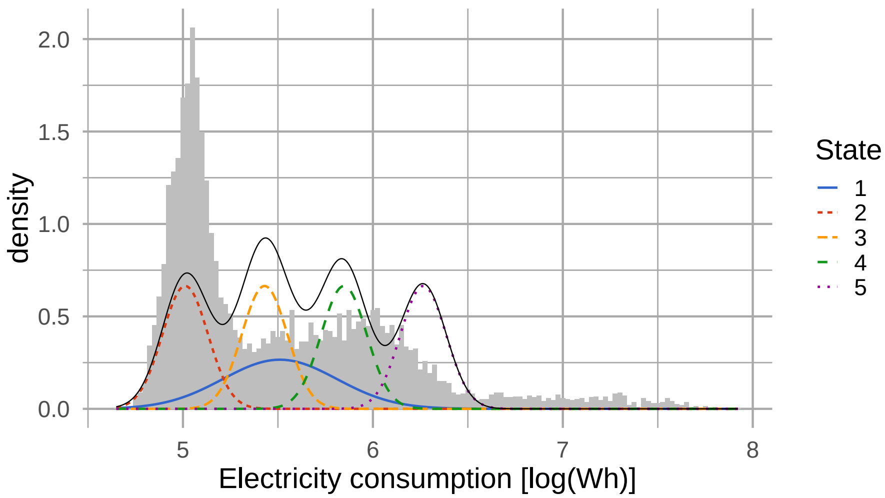


Figure D.1: Example of initial parameters for the state-dependent distributions for a 5-state model where state 1 is the initial transition state. Densities are shown with the histogram of the observations to indicate whether there are sufficient amounts of observations for each state-dependent distribution. The black line indicates the stationary distribution of the initial parameter setting. In this case, the probabilities of starting in each state are equal.

The transition probabilities are initialised such that $\hat{\gamma}_{ii} = 1 - m \cdot b$ and $\hat{\gamma}_{ij} = b$ for $i \neq j$ and suitable small b. Then there will be equal probability in the stationary distribution of starting in any state, which can be seen in Figure D.1.

An initial estimate of the transition rate q_i can be calculated from the discrete transition probability γ_{ii} by

$$\hat{q}_i = -\log \hat{\gamma}_{ii}. \tag{D.16}$$

This does not take into account the possibility of changing from a given state and back within an hour, hence this will be an underestimate of q_i, but for $q_i \ll 1$ the error is negligible [18]. Initialising w_i was done such that the intensity rates for shifting to the neighbouring states are equal, i.e., $w_i = 0.5$.

D.2.5 Pseudo residuals

Beside the use of information criteria to find a suitable model, we need to assess whether the model is adequate by assessing the goodness of fit, and we need to detect outliers relative to the model. One way to do this is to use pseudo-residuals [19].

For a continuous random variable X with continuous cdf, the cdf is uniformly distributed, i.e., $F(X) \sim U(0,1)$. Let Φ be the standard normal cdf, then $Z \equiv \Phi^{-1}(F(X))$ is standard normal distributed.

In this paper, we will use the pseudo-residuals that are based on the conditional distribution given all other observations under the model $(Pr(X_t = x_t|\mathbf{X}^{(-t)} = \mathbf{x}^{(-t)}))$, denoted ordinary pseudo-residuals [19]. For continuous observations, the ordinary pseudo-residuals are defined as

$$z_t = \Phi^{-1}\left(Pr(X_t \leq x_t|\mathbf{X}^{(-t)} = \mathbf{x}^{(-t)})\right). \tag{D.17}$$

The ordinary pseudo-residuals are used for validation of the model and for outlier detection. If the model is valid, the ordinary pseudo-residuals should be standard normal distributed. It should be noted that the pseudo-residuals might not be independent. It should be noted that the independence of the pseudo-residuals is not a requirement for the model validation in this paper.

D.2.6 Global decoding

Given a HMM and observations, the most likely state for a given time t can be determined by looking at the probability of observing this realisation for each state and choosing the one with the highest probability. This is called local decoding.

Global decoding is the determination of the most likely sequence of states conditioned on all the observations, i.e., now the probability of transition from the previous most likely state is also considered for $t \in \{1, ..., T\}$. This sequence is obtained by maximizing the conditional probability

$$Pr(\mathbf{C}^{(T)} = \mathbf{c}^{(T)}|\mathbf{X}^{(T)} = \mathbf{x}^{(T)}). \tag{D.18}$$

Loosely speaking, Equation (D.18) is the likelihood of $\mathbf{c}^{(T)}$ given all observations and model. Hence $\mathbf{c}^{(T)}$ can be used to identify model deficiencies and investigate the nature of the hidden states under the given model.

Global decoding is the key inference tool in this paper used for retrospective analysis. By grouping the sequence of states in accordance to time of day, we obtain a distribution of observed states for each hour of the day. The proportion of observations of a state for a given hour of the day is then an indication of the probability of observing this state in the given hour. Combining the observed proportions for all hours of the day we obtain a daily proportion profile for each state. With these profiles we get an indication of the most likely state for a given hour of day without fitting an inhomogeneous CT-HMM. Using the daily proportion profile in combination with the knowledge of sojourn times, state distributions, transition intensity matrix and the transition probability matrix, from the CT-HMM we will be able to infer knowledge of the states in relation to the underlying activities leading to the observed electricity consumption.

D.3 Data

The data used in this paper consists of time-series of electricity use at appliance level in one Swedish household. The data was obtained from a monitoring campaign carried out by the Swedish Energy Agency between August 2005 and December 2008 [22, 23] and has previously been studied in [24]. Most of these households where situated near Stockholm. The household chosen as the test case for this paper is an apartment with a family of two adults and a baby, with data covering one year. The electricity consumption data was measured in Wh with a precision of one decimal and a temporal resolution of 10 minutes. This data was aggregated to hourly time resolution, both at appliance level and for the total electricity use. This was to reflect the most common resolution data obtainable (total electricity purchased with hourly resolution) from the Danish DataHub [5]. Towards the end of 2020, a similar supplier-centric model DataHub will be available in Sweden [8], although the resolution of the data is not known yet.

Furthermore the studied household also kept a time-use diary, and this was used as validation/comparison data in this study. This household and its time-use diary have previously been analysed in detail as part of a behavioural study [25]. The total consumption was cleaned for zero observations and long periods of absence (Figure D.2), since long periods of absence do not tell anything about the diurnal patterns of electricity use in the household and introduce a risk of obtaining absorbing states in the Markov chain.

For this paper, the appliance time-series of the test case household has been regrouped such that a visible representation of these is possible (Figure D.3). The regrouping was done as follows:

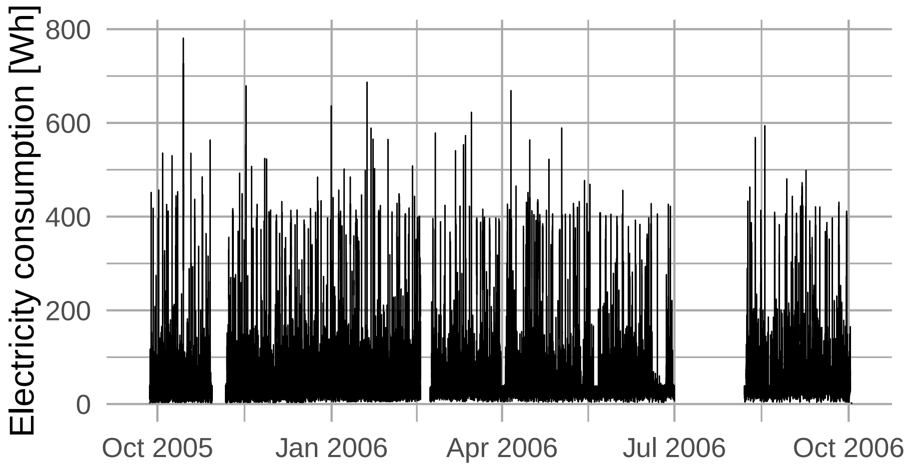


Figure D.2: The total electricity consumption for the studied household with hourly resolution.

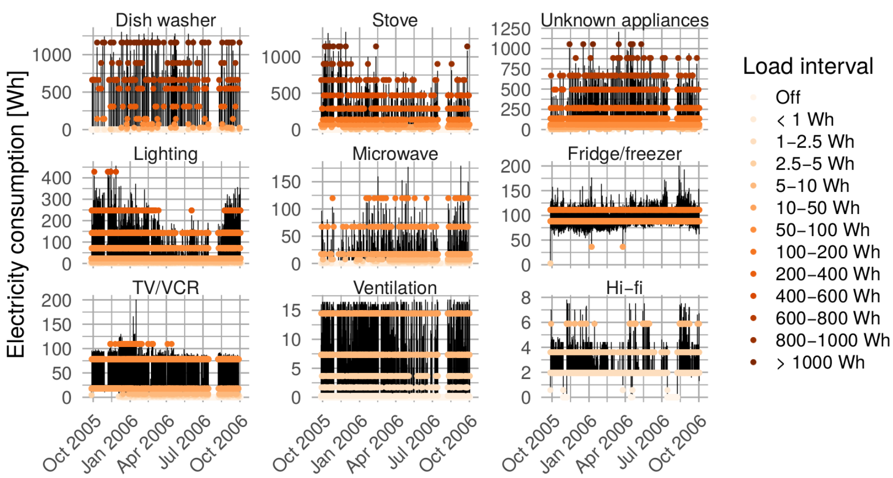


Figure D.3: The appliance time-series, shown with each observations load interval.

- Dishwasher, the individual load of the dishwasher.
- Stove, the cumulated load of all stove phases.

- Unknown appliances, appliances not metered individually and appliances which were only metered for part of the period.
- Lighting, the cumulated lighting load of the whole household.
- Microwave, the individual load of the microwave.
- Fridge/freezer, the cumulated load of the fridge and freezer.
- TV/VCR, the cumulated load of the TV and VCR.
- Ventilation, the individual load of the ventilation system.
- Hi-Fi, the individual load of the Hi-Fi system.

To investigate which appliances are used and at what intensity at a given time, each observation of the appliance time-series was categorized according to load intervals. These intervals were chosen to provide enough intervals to distinguish between different load intensities and standby use for all appliance types (Figure D.3). E.g., comparing TV/VCR with Hi-Fi, the standby load observed differs in level. For the lighting and ventilation there is also a wide span of intensities in the load intervals. This means that the load intervals can only be interpreted in relation to an individual appliance time-series and not in general. These load intervals are not used in the model fitting, but they are applied in the analysis of the states in the CT-HMM after these have been identified. For a given state and time of day, we can obtain the proportion of load intervals for each appliance time-series, and with this we have an indication of the appliance-related activities in the household.

D.4 Results

Table D.1: The transition intensity matrix $\boldsymbol{Q}$.

	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
State 1	-0.955	0.208	0.000	0.000	0.000	0.000	0.000	0.000	0.747
State 2	0.337	-0.571	0.234	0.000	0.000	0.000	0.000	0.000	0.000
State 3	0.000	0.620	-0.774	0.154	0.000	0.000	0.000	0.000	0.000
State 4	0.000	0.000	0.366	-0.422	0.056	0.000	0.000	0.000	0.000
State 5	0.000	0.000	0.000	0.443	-0.473	0.031	0.000	0.000	0.000
State 6	0.000	0.000	0.000	0.000	2.569	-3.030	0.461	0.000	0.000
State 7	0.000	0.000	0.000	0.000	0.000	1.963	-2.171	0.208	0.000
State 8	0.000	0.000	0.000	0.000	0.000	0.000	1.001	-1.425	0.424
State 9	0.157	0.000	0.000	0.000	0.000	0.000	0.000	0.618	-0.775

A series of models with 5 to 12 states were fitted to the log-transformed total electricity consumption shown in Figure D.2 and with Gaussian state distribution. Based on the results of the information criteria and the optimization results in Table D.3, the 9-state model was chosen as the most suitable, both for

Table D.2: The transition probability matrix $\boldsymbol{P}$, for 1 hour time step. The most likely state transitions are underlined, for each state these will approximately sum up to 91-96% of the transition probability.

	State 1	State 2	State 3	State 4	State 5	State 6	State 7	State 8	State 9
State 1	$\underline{0.426}$	$\underline{0.103}$	0.012	0.001	0.005	0.008	0.022	$\underline{0.086}$	$\underline{0.338}$
State 2	$\underline{0.167}$	$\underline{0.622}$	$\underline{0.125}$	0.010	0.001	0.001	0.002	0.011	0.061
State 3	0.050	$\underline{0.332}$	$\underline{0.513}$	$\underline{0.088}$	0.003	0.000	0.000	0.002	0.012
State 4	0.007	$\underline{0.065}$	$\underline{0.210}$	$\underline{0.681}$	0.036	0.000	0.000	0.000	0.001
State 5	0.001	0.010	0.048	$\underline{0.289}$	$\underline{0.644}$	0.008	0.001	0.000	0.000
State 6	0.000	0.004	0.025	$\underline{0.191}$	$\underline{0.641}$	$\underline{0.087}$	0.044	0.006	0.001
State 7	0.001	0.001	0.010	$\underline{0.096}$	$\underline{0.477}$	$\underline{0.186}$	$\underline{0.175}$	0.043	0.012
State 8	0.012	0.001	0.002	0.023	$\underline{0.166}$	$\underline{0.129}$	$\underline{0.205}$	$\underline{0.304}$	$\underline{0.157}$
State 9	$\underline{0.071}$	0.008	0.001	0.003	0.030	0.037	$\underline{0.083}$	$\underline{0.229}$	$\underline{0.538}$

simplicity and since we cannot reject that the 10-state model has local optimum or is stuck in the optimization.

Table D.3: The information criteria results of the model fits. The code describes the optimization result from the `nlm`-function in `R`, as described in Section D.2.

no. states	AIC	BIC	code
5	6115.55	6254.54	1
6	6023.55	6190.33	2
7	5701.52	5896.10	1
8	4956.20	5178.58	1
9	4851.55	5101.72	1
10	4788.87	5066.84	3
11	5962.49	6268.25	3
12	6088.94	6422.51	3

To assess the goodness of fit of the 9-state model, we look at the ordinary pseudo-residuals (Figure D.4) to check whether there is a significant deviation from the observations. From the histogram and the qq-plot, a slight skewness of the pseudo-residuals is observed. This indicates that the log transform of the data might not be the optimal, and some other transformation of the data could be considered. In [24], the best fit was not log-normal, but Weibull, for similar time-series. In this setting the slight skewness of the pseudo-residuals is acceptable, since we only do retrospective analysis and not forecasting, and the model chosen is the 9-state model.

To get an overview of the underlying distributions of each state in the CT-HMM, the stationary distribution is investigated (Figure D.5 and Table D.4). The stationary distribution is calculated based on the 1-hour step transition probability matrix (Table D.2), which is calculated from Table D.1 (see Section D.2). From this, a semi-periodic sequence between State 1, with the highest mean value in the underlying distribution, to State 4, with the lowest mean value in the underlying distribution, is observed. This observation is based on which other state has the highest probability of being next for a given state.

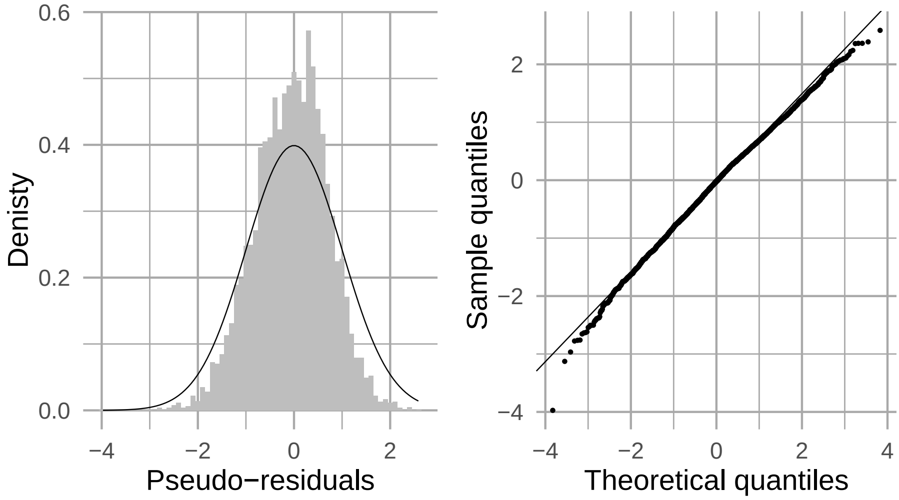


Figure D.4: Histogram and qq-plot of the ordinary pseudo-residuals for the 9-state CT-HMM.

This periodic sequence or cycle is $(1, 9, 8, 7, 6, 5, 4, 3, 2)$, since if starting in State 1, the most likely state to enter next is State 9 etc. as seen in both the transition intensity- and probability matrix in Tables D.1 and D.2.

Table D.4: The parameter estimates of the state distributions, where μ is the mean and σ the standard deviation of the Gaussian distributions. Note that the estimates are based on the log-transformed data.

State	$\hat{\mu}$	$\hat{\sigma}$
1	6.82	0.48
2	5.66	0.26
3	5.13	0.10
4	4.95	0.08
5	5.05	0.06
6	5.26	0.09
7	5.52	0.10
8	5.85	0.12
9	6.12	0.15

Looking at the sojourn times described in Section D.2, and shown in Figure D.6, we also learn something about the dynamics of the states. If the sojourn time is less than an hour, these states occur rarely and only once before changing to a different state. This indicates that these states are describing transitions between activities, e.g., the dishwasher cycle ends at the start of the cumulated observation of electricity use the hour where one of these states are found most

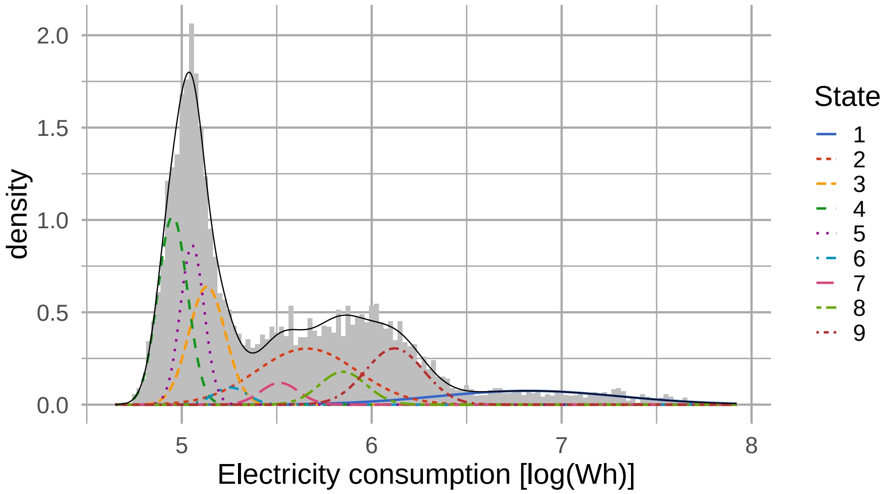


Figure D.5: The stationary distribution of the CT-HMM and the underlying Gaussian distributions of the states scaled by the stationary distribution.

likely.

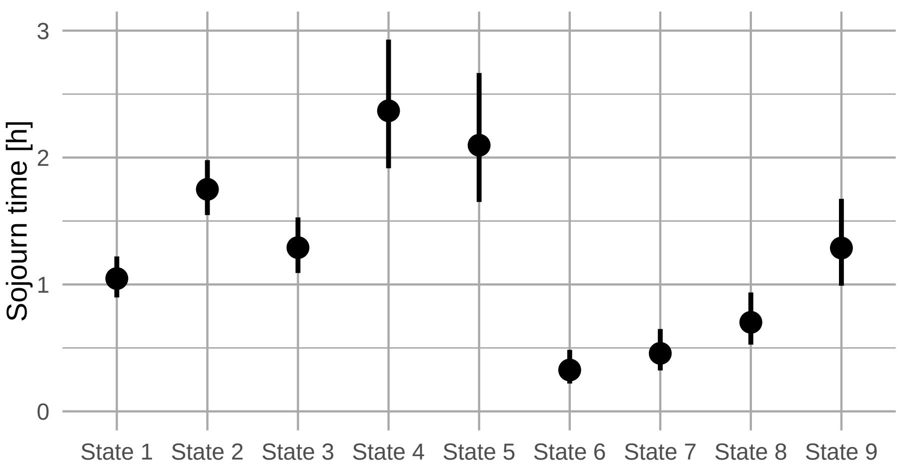


Figure D.6: The estimated sojourn times of the CT-Markov chain, with 95% confidence band.

For states with sojourn times greater than 2 hours, we have indication that these states often are observed multiple times following each other and that the probability of visiting other states between the observations is low. This indicates longer periods of stable electricity consumption, hence low interaction with appliances.

For states with sojourn times close to 1, we have indications of more interaction with appliances, since there is a more equal probability of staying or switching state in the next observation.

D.4.1 Global decoding

The most likely sequence of states (Section D.2) was determined based on all the data shown in Figure D.2, but in the following we only consider the sequence from weekdays, i.e., excluding weekends (Figure D.7). This is due to the difference observed in the diurnal pattern between weekdays and weekends. Weekends are not presented in this paper since the main difference is a shift in time in the proportion profiles, since the family seems to wake up later on weekends compared to weekdays.

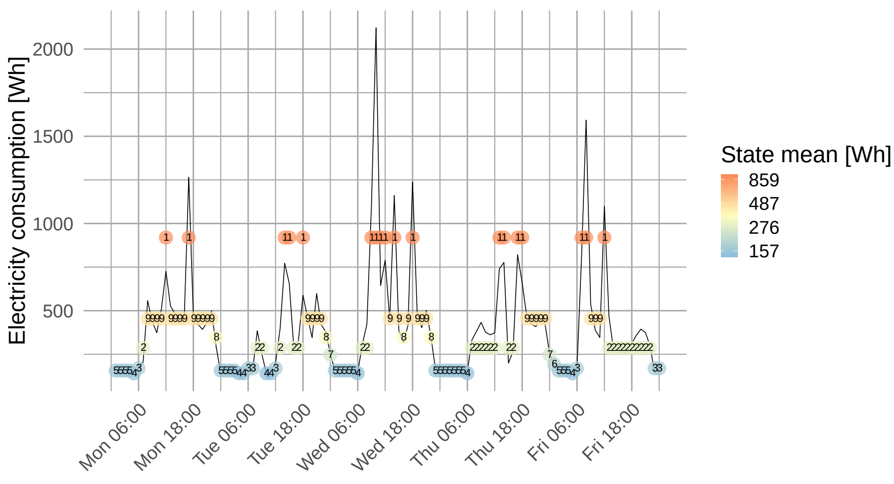


Figure D.7: Subset of the most likely sequence of states from 4^{th} of September to 9^{th} of September 2006.

From the proportion profiles (Figure D.8), the proportion of state occurrences at a given time of day is seen for each state. These profiles indicate the probabilities for each state at a given hour and can be linked to the possible underlying activities. Comparing Table D.1 and Figures D.7 and D.8, we see a clear daily pattern in the cycle through the states. Starting at midnight in State 5 shifting through States 4-2 during the day and to States 1 and 9 in the evening ending back at State 5 around midnight through States 6-8. This daily pattern suggests that the states are time dependent and an inhomogeneous CT-HMM could be applied.

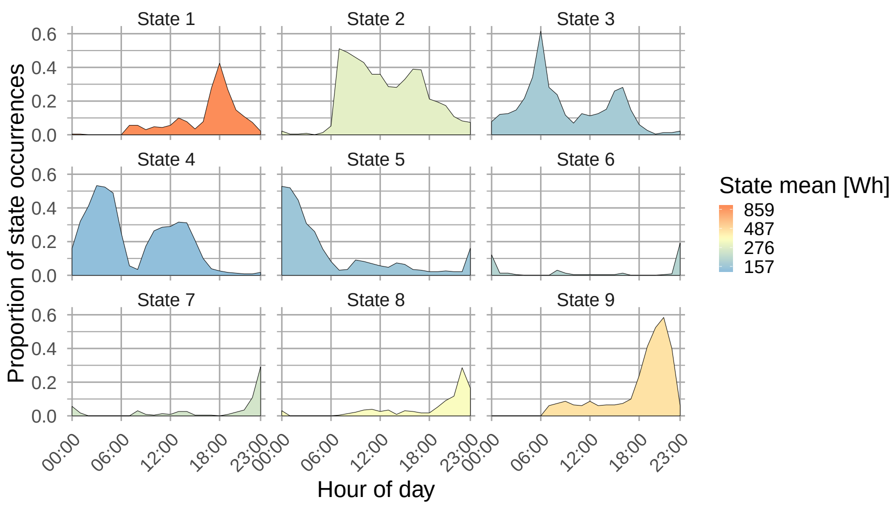


Figure D.8: The proportion of state occurrences given time of day, coloured by the estimated mean value of the underlying state distribution.

D.4.2 State descriptions

By investigating the underlying appliance loads for each state, together with the proportion profiles, sojourn times, transition intensities, transition probabilities and the underlying state distributions, the following state descriptions in accordance with appliance-related use activity were found reasonable:

- State 5: Minimum activity, high base load or minor night time activity.
- State 4: Minimum activity, low base load.
- State 3: Ramp up transition state, high base load or minor night time activity.

- State 2: Medium load daytime activity (TV, cooking, medium lighting load)
- State 1: Multiple and/or high load activity (dishwasher, cooking, vacuum e.t.c.).
- State 9: Medium/high load evening activity, (TV, cooking, high lighting load).
- State 8, 7 and 6: Ramp down transition states.

In the remainder of this section, the reasoning for the above descriptions is presented in detail. We will go through the states in the observed daily cycle, starting with the low mean States 4 and 5. At the end of this section, the state descriptions are compared to a time-use diary reported by the residents over a period of four days.

It should be noted that in the following, Figures D.9-D.14 are based on Figures D.3 and D.8. For each state, the proportion profiles from Figure D.8 are presented with the underlying observed loads for each appliance time-series from Figure D.3.

D.4.2.1 States 4 and 5

For States 4 and 5, the underlying distributions have the lowest mean values of all the states (Table D.4), which indicates a high probability of appliance inactivity. These states could be described as base load states, where the base load is defined as standby use plus the appliances that turn on and off within an hour without any interaction, e.g., fridges and freezers. The long sojourn times also hint at inactivity since it is expected that sleep and absence are observed over a continuous period. From the proportion profile for States 4 and 5 in Figure D.8, it is seen that there is a high proportion of occurrences for these states at night, for State 4 this is also the case during the day, which confirms the previous findings.

To examine the differences between States 4 and 5, first the underlying appliance loads are investigated. For State 4 there are in general low load levels relative to the individual appliances (Figure D.9). The small peaks in lighting, ventilation and microwave at 06:00 and 18:00 correspond well with the decrease in occurrence proportion for this state (Figure D.8). This indicates that, e.g., the light has just been turned on for a very short time within the observed hour, hence the next state observed is most likely to be an activity state. The observed loads for the TV/VCR seem to be standby use when not off (see Figure D.3).

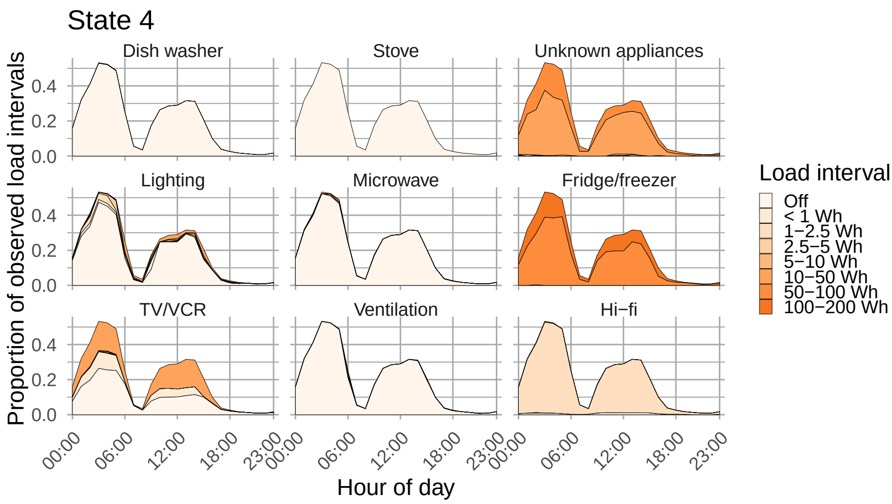


Figure D.9: The proportion of occurrences for State 4 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

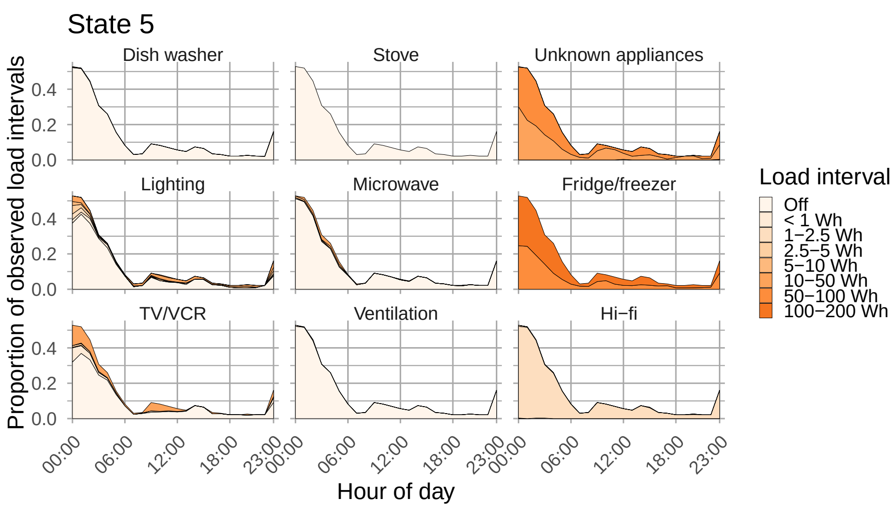


Figure D.10: The proportion of occurrences for State 5 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

For State 5 (Figure D.10), almost the same as for State 4 is observed, except that the fridge/freezer and the unknown appliances have higher proportions of slightly higher loads. It is observed for lighting and microwave that there is a slightly higher proportion of a small load during the night compared to State 4, this could indicate that State 5 could also be observed in hours with minor night time activities, e.g., turning on the light when going to the toilet, or heating food for the baby. It could also be due to ramping down from a higher consumption, e.g., the lights have been on for a short time and then turned off.

The last thing to consider are the transition intensities and probabilities (Tables D.1 and D.2). These can tell how the state will change in the future. Given that the electricity consumption is in State 5, there is a high probability of staying in this state or shifting to State 4 and low probability of shifting to State 6.

D.4.2.2 State 3

Looking at the proportion profile for State 3 in Figure D.8, there is a relatively high proportion during the night compared to States 1 and 2, followed by a peak at 6:00 and again at 15:00-16:00. After the second peak, the proportion becomes very low. Investigating the estimated sojourn time, transition intensities and probabilities, there is a high probability of staying in the state or of changing to State 2. There is also some probability of changing back to State 4.

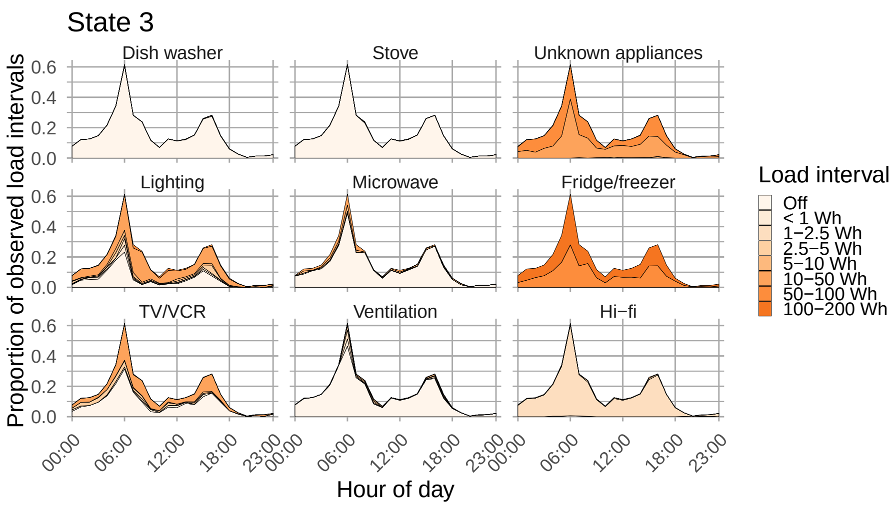


Figure D.11: The proportion of occurrences for State 3 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

The underlying appliance loads are investigated from Figure D.11. Compared to State 4, there are higher proportions for the higher loads in both fridge/freezer and unknown appliances, as seen for State 5. There does not seem to be any significant difference when looking at Hi-Fi and TV/VCR. The main difference is in lighting, microwave and ventilation. First it is observed that a higher proportion of the lighting is turned on. The proportion of microwave use is similar to State 5 at night. For ventilation there is higher proportion of low loads in the morning and afternoon, which indicates the beginning of activity in the kitchen.

Based on the proportion profile and the underlying appliance loads, State 3 seems to be related to starting activities, when the residents wake up in the morning or are coming home from work, but also some minor activity during night time.

D.4.2.3 State 2

The underlying distribution (Figure D.5 and Table D.4), has a high variance compared to all other states except for State 1. This indicates a considerable variation in combinations of appliance use. Looking at the proportion profile (Figure D.8), low proportions for this state during night are seen, with a great increase in the morning, which then slightly decreases over the day with a minor peak in the late afternoon. Looking at the sojourn time, transition intensities and probabilities, there is a high probability of staying in this state for consecutive time steps. These observations indicate daytime activities.

When investigating the underlying appliance loads (Figure D.12), it is observed that there is a high proportion of TV/VCR and Hi-Fi use compared to the previously described states. For the unknown appliances, two new load intervals covering a third of the proportion of the observed load levels are seen. This indicates medium load appliances like computers, kitchen machines, etc. Looking at the stove, microwave and ventilation, an increase in the proportion of use of these is observed. Furthermore, there seems to be a correlation between the stove and ventilation in the morning, and the microwave and ventilation in the evening, but not at midday. This could indicate less load-intense cooking activity. A high increase in the lighting load is also observed.

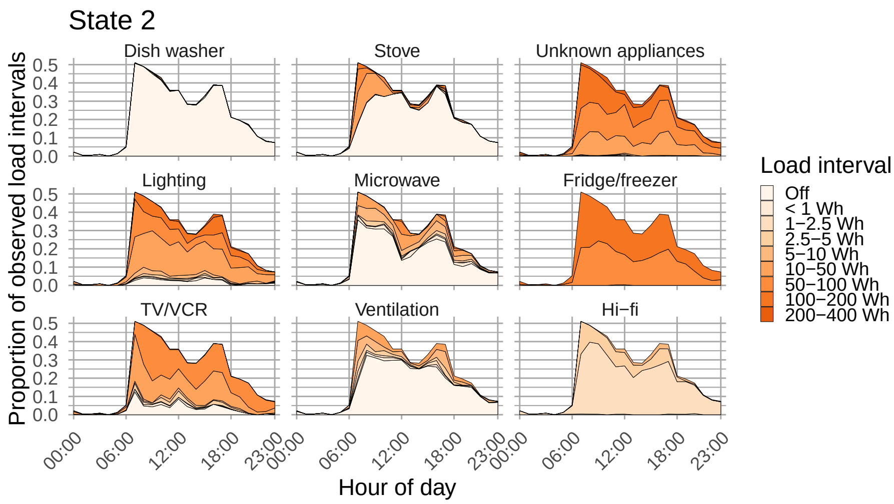


Figure D.12: The proportion of occurrences for State 2 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

D.4.2.4 State 1

From the underlying distribution (Figure D.5 and Table D.4), we have a high mean value and a high deviation, hence the combination of appliances used must have a high load. Whether the high load is due to one appliance or a combination is not known. Looking at the proportion profile (Figure D.8), it is seen that there are no observations of this state during the night, stable proportion until the afternoon where a peak is seen at 18:00, then a slow decrease towards midnight. From the sojourn time, transition intensities and probabilities, this state will often be observed only once or twice in a row before switching state again. Depending on the time of day, the observed switch will be either to State 2 or State 9.

Looking at the underlying appliance loads (Figure D.13), the unknown appliances and the lighting have very high load compared to the rest of the states. There is also high proportion of use of the TV/VCR and the Hi-Fi. This indicates a broad range of activities taking place at the same time or using high load appliances like vacuum cleaners, hair-dryers or water kettles. Looking at the stove, microwave and ventilation, a high proportion of use of the stove in the morning and then from midday to the evening is observed. The load levels in the morning are lower than midday and evening, which indicates shorter use

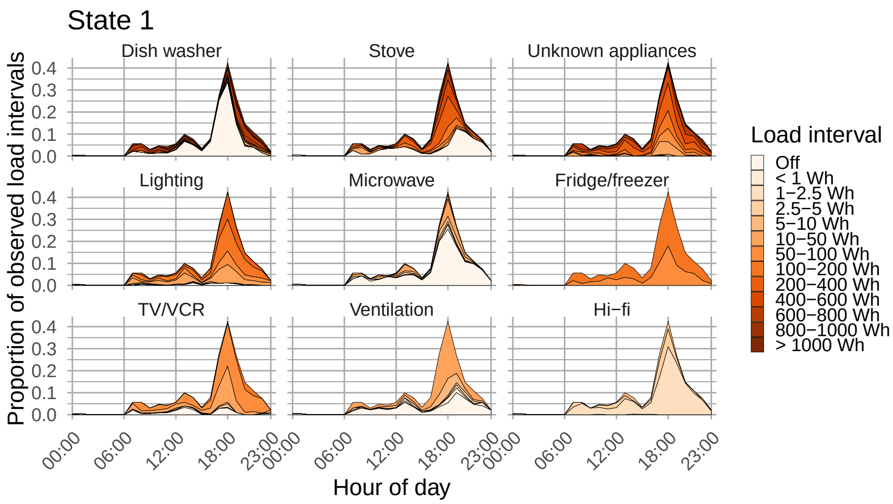


Figure D.13: The proportion of occurrences for State 1 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

of the stove in the morning. There seems to be a correlation between the use of the stove and ventilation. The microwave seems to have a similar profile as for State 2. For the dishwasher, it is observed that this appliance is only used in State 1, and mainly around the peak in the proportion profile. Furthermore, it is noted that observing the dishwasher and the stove having a high load at the same time occurs seldom.

D.4.2.5 State 9

There is a similar proportion profile (Figure D.8) for State 9 as for State 1, except that the peak shifts towards midnight, hence there is a high proportion of this state in the evening. The sojourn time is similar to the sojourn times of States 2 and 3, which indicates high probability of successive observations of State 9. Looking at the transition intensities and probabilities, there is higher probability of shifting to State 8 than to State 1.

Looking at the underlying appliance loads (Figure D.14), the unknown appliances and the lighting have very high load compared to the rest of the states, except for State 1. There is also high proportion of use of the TV/VCR and the Hi-Fi. There seems to be a correlation between the use of the stove and

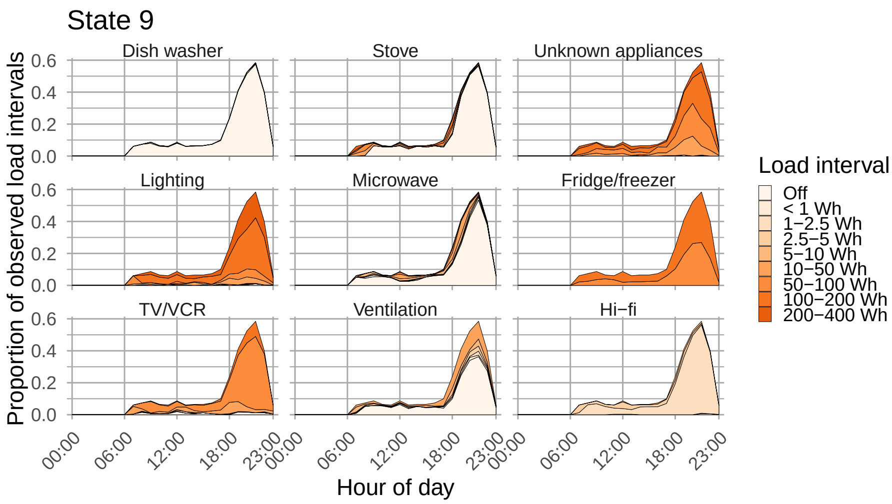


Figure D.14: The proportion of occurrences for State 9 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

ventilation. Given the high transition probability of going from State 1 to State 9, the stove activity could be due to ramping down from a higher load. The microwave seems to have a similar profile as for States 1 and 2, but with more visible peaks.

D.4.2.6 States 8, 7 and 6

The estimated sojourn times, which are less then 1 hour, and the very low transition intensities tell us that these states are occurring rarely, and only once before changing to a different state. From the proportion profiles (Figure D.8), it is seen that in general these states are observed rarely, except just before and around midnight. This again reflects the transition from being active to inactive. It should be noted that the distributions of States 6, 7 and 8 are all contained in the distribution of State 2 (Figure D.5). State 8 is a bit different when looking at the transition probabilities, since there is some probability of going back to State 9.

D.4.2.7 Comparison to time use diary

To validate the description of the states, a comparison with a time-use diary from the household is carried out. The diary covers 4 days over a weekend from Friday to Monday. Recall that the observed electricity consumption at a given hour is the cumulated consumption within the previous hour.

From Figure D.15, a clear correspondence between State 1 and longer cooking activities and turning on the dishwasher is observed. The dishwasher activity covers only the short time it takes to turn it on, hence it is difficult to see in the plot, but the male resident turns it on Friday 07:10 and again Sunday 09:40.

States 3, 4 and 5 are observed when the residents are sleeping or absent. It should be noted that the work activity can be differentiated into home or elsewhere. If there is an away activity before and after a work period it is elsewhere, otherwise home.

States 2 and 9 are observed when there are many different activities within the hour of the observed consumption, including shorter cooking activities or activities covering a longer period like watching TV or working at home. It is also observed that State 9 is more likely to occur when both residents are at home doing different things. This might indicate that State 2 is more likely

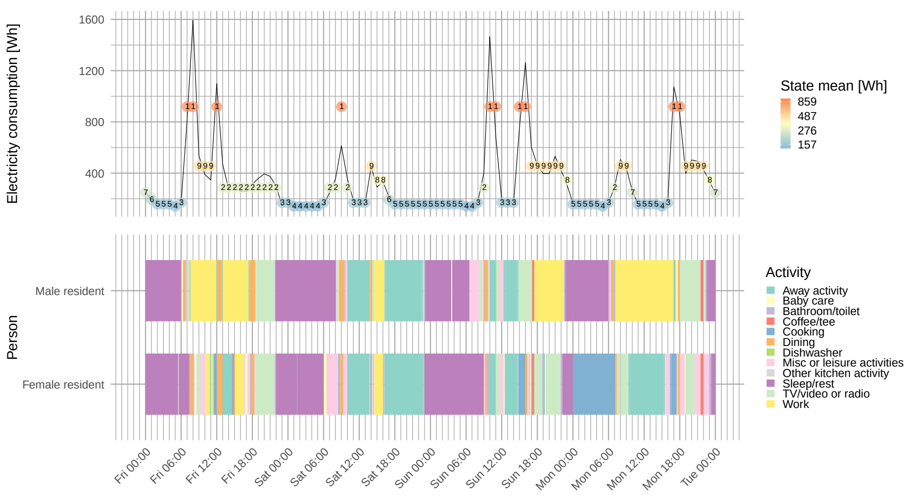


Figure D.15: Comparison of the most likely sequence of states with time-use diary.

when only one adult resident is at home or a common activity is going on, such as watching TV.

States 6, 7 and 8 are observed when transition from an active to an inactive state is occurring within the hour of observed consumption, which corresponds well to the previous findings.

The state descriptions are well aligned with the time diary, although a ramp down through States 2 and 3 is observed. This is not an unexpected event due to the nature of the CT-HMM, but the ramp down is more likely through States 6, 7 and 8.

D.5 Concluding discussion

We have successfully applied an approach for disaggregation of electricity consumption in a single household using CT-HMM. By applying CT-HMM for hourly observations of total electricity consumption of the household, we have described the obtained states in accordance with appliance-related activities using global decoding and a comparison to specific appliance loads. The parameter reduction obtained using the CT-HMM have greatly reduced the dimension of the solution space in the optimization compared to a discrete-time HMM with the same number of states.

Given multiple persons acting at the same time, and due to the measurement, resolution is on hourly basis, each observation can be assumed to be a realisation of a mixture of activities. The activities can spread over several observations and can have variations in duration, e.g., there is a huge difference in energy consumption between boiling an egg or a kilo of potatoes, but the activity is the same. Hence the states cannot be described in direct accordance with appliances, nor in accordance with a specific activity, however we can describe the states in accordance with sets of appliance-related activities and levels. The state descriptions were compared to a time-use diary and we found a good correspondence between reported activities and the state descriptions.

The results of this study are specific to the household studied and further studies are needed to investigate whether the state descriptions can be generalized or whether there are more types of states. It should be noted that using this initial model framework on households with electric heating or cooling will not yield useful results due to the dynamics of the temperature influence. Either the temperature-dependent consumption should be disaggregated and this framework can be utilised on the remaining consumption, or the temperature should enter the model framework as covariates in the state-dependent distributions. Further studies are needed to determine whether such integration of the temperature will yield the desired descriptions of the states in accordance with appliance activities.

In a future case of a proportion profile using data from the DataHub, we do not have direct access to appliance-specific data, and an explicit analysis of what is going on in the states is impossible. With no appliance-specific data, the states could instead be described in accordance to their distribution, sojourn time, average probability profile and transition intensities/probabilities. Assuming these state descriptions hold for households in general, knowledge about high-load appliances in a household might narrow the possible set of activities, which the consumer can compare with his/her activity.

Further investigation into how to initialize the parameters for the optimization in order to make a general set-up capable of handling many types of consumption patterns is needed. The approach used in this paper is partly based on a subjective method. In principle, there exist m solutions to the optimization problem, e.g., a solution where State 1 is the high activity state (highest mean value) but also one where State 5 is the high activity state. The solution found depends highly on the initialisation, hence best practises to avoid starting between two such solutions are needed to avoid local minima and reduce computation time.

Applications for consumers with the current state of the model framework are to do retrospective summaries of their consumption, how much has been used in each state, how much time has been spent in each state and when? Such information might help consumers to relate their actions on appliance use to the load and yield insights otherwise overlooked.

Further model development could take the daily and/or weekly time dependence into account, i.e., create an inhomogeneous Markov Model and/or to include explanatory variables like the temperature. This could improve the model framework's forecasting properties, such that it could be used for residential demand-side management for cost-load optimization.

Acknowledgements

This work is partly funded by Innovation Fund Denmark (IFD) under File No. 5189-00045B and Centre for IT-Intelligent Energy Systems in Cities (CITIES) project (DSF 1305-00027B).

References D

[1] K. C. Armel et al. "Is disaggregation the holy grail of energy efficiency? The case of electricity". In: *Energy Policy* 52 (2013), pp. 213–234.

[2] B. Najafi, S. Moaveninejad, and F. Rinaldi. "Data analytics for energy disaggregation: methods and applications". In: *Big Data Application in Power Systems*. Elsevier, 2018, pp. 377–408.

[3] H. Madsen. *Time series analysis*. CRC Press, 2007.

[4] J. Kelly and W. Knottenbelt. "Does disaggregated electricity feedback reduce domestic electricity consumption? A systematic review of the literature". In: *arXiv preprint arXiv:1605.00962* (2016).

[5] *What is DataHub?* `https://en.energinet.dk:443/Electricity/DataHub`. Accessed:2018-02-13. 2018.

[6] *bekendtgørelse om fjernaflæste elmålere og måling af elektricitet i slutforbruget*. `https://www.retsinformation.dk/Forms/R0710.aspx?id=160434`. Accessed: 2019-04-16. 2013.

[7] *Smart Metering deployment in the European Union*. `https://ses.jrc.ec.europa.eu/smart-metering-deployment-european-union`. Accessed: 2019-04-17. 2019.

[8] *Data hub*. `https://www.svk.se/en/stakeholder-portal/Electricity-market/data-hub/`. Accessed: 2018-02-13. 2018.

[9] *2012/148/EU: Commission Recommendation of 9 March 2012 on preparations for the roll-out of smart metering systems*. `https://eur-lex.europa.eu/eli/reco/2012/148/oj`. Accessed: 2019-04-17. 2012.

[10] G. W. Hart. *Nonintrusive Appliance Load Data Acquisition Method: Progress Report*. MIT Energy Laboratory, 1984.

[11] G. W. Hart. "Nonintrusive appliance load monitoring". In: *Proceedings of the IEEE* 80.12 (1992), pp. 1870–1891.

[12] J. Z. Kolter, S. Batra, and A. Y. Ng. "Energy disaggregation via discriminative sparse coding". In: *Advances in Neural Information Processing Systems*. 2010, pp. 1153–1161.

[13] B. Zhao, L. Stankovic, and V. Stankovic. "Electricity usage profile disaggregation of hourly smart meter data". In: *4th International Workshop on Non-Intrusive Load Monitoring*. 2018.

[14] B. J. Birt et al. "Disaggregating categories of electrical energy end-use from whole-house hourly data". In: *Energy and Buildings* 50 (2012), pp. 93–102.

[15] A. Kipping and E. Trømborg. “Modeling and disaggregating hourly electricity consumption in Norwegian dwellings based on smart meter data”. In: *Energy and Buildings* 118 (2016), pp. 350–369.

[16] J. Liisberg et al. “Hidden Markov Models for indirect classification of occupant behaviour”. In: *Sustainable Cities and Society* 27 (2016), pp. 83–98.

[17] P. Nystrup, H. Madsen, and E. Lindström. “Stylised facts of financial time series and hidden Markov models in continuous time”. In: *Quantitative Finance* 15.9 (2015), pp. 1531–1541.

[18] H. Madsen, H. Spliid, and P. Thyregod. “Markov models in discrete and continuous time for hourly observations of cloud cover”. In: *Journal of climate and applied meteorology* 24.7 (1985), pp. 629–639.

[19] W. Zucchini and I. L. MacDonald. *Hidden Markov models for time series: an introduction using R*. CRC Press, 2009.

[20] M. Pinsky and S. Karlin. *An introduction to stochastic modeling*. Academic press, 2010.

[21] C. H. Jackson et al. “Multi-state models for panel data: the msm package for R”. In: *Journal of Statistical Software* 38.8 (2011), pp. 1–29.

[22] J. P. Zimmermann. “End-use metering campaign in 400 households in sweden assessment of the potential electricity savings”. In: *Contract* 17.September (2009), pp. 5–2743.

[23] *Elmätning i bostäder*. `http://www.energimyndigheten.se/statistik/bostader-och-lokaler/forbattrad-energistatistik-i-bebyggelsen-och-industrin/elmatning-i-bostader/`. Accessed: 2018-02-13. 2016.

[24] J. Munkhammar, J. Rydén, and J. Widén. “Characterizing probability density distributions for household electricity load profiles from high-resolution electricity use data”. In: *Applied Energy* 135 (2014), pp. 382–390.

[25] *hushållens elanvändningsmönster identifierade i vardagens aktiviteter*. `http://www.energimyndigheten.se/globalassets/statistik/festis/beteendestudier/hushallelanv2.pdf`. Accessed: 2018-07-09. 2008.

D.6 Appendix

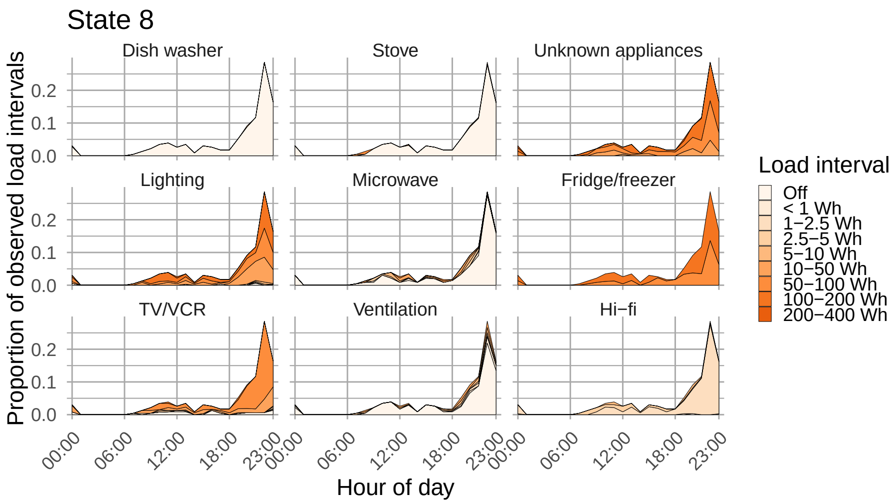


Figure D.16: The proportion of occurrences for State 8 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

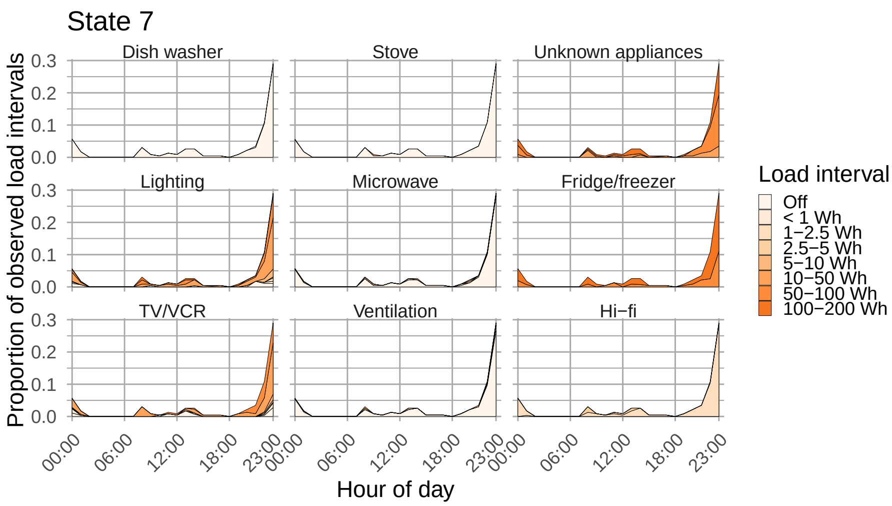


Figure D.17: The proportion of occurrences for State 7 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

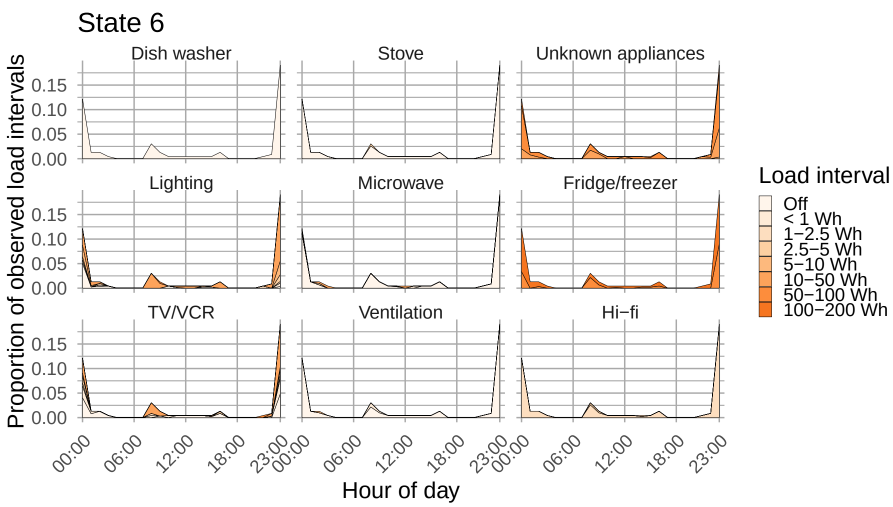


Figure D.18: The proportion of occurrences for State 6 given time of day, shown with the underlying proportions of observed loads for each appliance time-series.

Paper E

Gamification of electricity consumption, PowerMon

Authors:
Jon Liisberg

Technical report:
DTU library.

Gamification of electricity consumption, PowerMon

Jon Liisberg[1,2]

Abstract

By setting up a definition of gamification a description of gamification in relation to electricity consumption is formulated. Based on this a game prototype is developed and presented. The game is tested and evaluated on users of the digital energy assistant app Watts. The players found it difficult to understand the game, leading to suggestion for improvement. Indication of some relation to their electricity consumption was observed through the game play.

E.1 Introduction

Beside providing feedback through the app Watts [1], based on data-driven models to the consumers to help them save electricity, engaging or entertaining approaches could also help the consumers become aware of their consumption patterns.

Several studies on gamification and serious games within the domain of residential energy consumption are reviewed in [2]. From this review it is concluded that in general, there is a positive effect on reducing energy consumption, but also that there is a limited amount of empirical evidence suggesting more rigorous follow-up studies are needed.

In the following, we will be looking at the approach of advice through gamification. For this it is needed to establish a notion of gamification in respect to electricity consumption and how this can be applied. To do this, definitions of games and game design are needed. With these definitions a goal for gamification of electricity consumption is formulated and based on this a prototype game

[1]DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark
[2]SEAS-NVE, Digital Innovation, Hovedgaden 36, DK-4520 Svinninge, Denmark

is developed. This prototype is play tested with four Watts users, to investigate if the game is playable and if the players see the relation between actions in the household and electricity consumption, through the game experience.

E.2 Games and game design

To clarify the meaning of games and design in relation to games, definitions of the two will be formulated as a base for describing gamification and to discuss gamification of electricity consumption.

First, we will look at the relation between play and games. One general definition of play is: "play is free movement within a more rigid structure" [3] (Ch. 22, p. 4). In [3] (Ch. 22) it is further elaborated that games are a subset of play and game play is the play arising from the structure of the game. In [3] a definition of games is given, this definition is based on a comparison study of eight definitions from a variety of fields and is formulated as follows.

"A game is a system in which players engage in an artificial conflict, defined by rules, that results in a quantifiable outcome" [3] (Ch. 7, p. 11).

We will stick to the above definition in the setting of this project. To specify the elements of this definition, these are summarized as follows.

- System: "A system is a set of parts that interrelate to form a complex whole." [3] (Ch. 5, p. 6) Within a system there are four elements that all systems share. Objects, attributes, internal relationship and environment.
  - *Players: A game is something that one or more participants actively play. Players interact with the system of a game in order to experience the play of the game.*
  - *Artificial: Games maintain a boundary from so-called "real life" in both time and space. Although games obviously occur within the real world, artificiality is one of their defining features.*
  - *Conflict: All games embody a contest of powers. The contest can take many forms, from cooperation to competition, from solo conflict with a game system to multiplayer social conflict. Conflict is central to games.*
  - *Rules: . . . Rules provide the structure out of which play emerges, by delimiting what the player can and cannot do.*
  - *Quantifiable outcome: Games have a quantifiable goal or outcome. At the conclusion of a game, a player has either won or lost or received some kind of numerical score. A quantifiable outcome is what usually distinguishes a game from less formal play activities. [3] (Ch. 7, p. 11)*

There are of cause many different definitions of what games and play are but with this definition and specification of the definition elements a reference point is established, when discussing gamification, game design and the game design approach used in this project.

Design is again an ambiguous concept depending on the setting and purpose. In this setting we have chosen the following definition. “game design is the process by which a designer creates a context to be encountered by a participant, from which meaning emerges.” [3] (Ch. 7, p. 11).

For this project an iterative design process is used, where prototyping and play-testing is emphasized. The iterative design process consists of four steps: prototyping, play-testing, evaluation and revision. Very early in the process a crude prototype of the game is made. This prototype is then play-tested and evaluated in relation to the game definition. With the evaluation a revision of the prototype is carried out and the game will be tested again. The play-testers will change during this cyclic process starting with self-testing moving on to confidants, ending with testers of the target audience of the game. For further reading see [3] and [4].

E.3 Gamification

In this section a definition of gamification is presented, and then applied to the setting of electricity consumption.

Gamification can be defined as: “using game design elements in a non-game context” [5]. The first question that arises is, what is game design elements? In [5] it is suggested that game design elements are restricted to “elements that are found in most games, readily associated with games, and found to play a significant role in game-play.” The second question is what is context?

The following quote is given to describe context “a frame that surrounds the event and provides resources for its appropriate interpretation.” [6]. This is a somewhat static description.

In [7] Dourish discuss context in the setting of human-computer interaction or ubiquitous computing where computer technology is integrated with the everyday physical world. Dourish argues that context is not a fixed description of a setting “Context cannot be a stable, external description of the setting in which activity arises. Instead, it arises from and is sustained by activity itself”. Dourish concludes that context is a slippery notion. We could state: Context are actions that arise and are rendered meaningful in the human-computer interaction.

In [8], Deterding argues that the context is very important in game design. Game design should be developed based on the context and not using a predefined game design and force it over the context, which have been seen numerous times like scoring points and gaining badges.

In order to describe gamification related to electricity consumption we need a problem description in order to formulate the goal.

To a large extend many people today, have no clear idea about how they are using electricity. And it is very difficult to give direct advice about e.g. appliances that are turned on and off based only on hourly measures of electricity use. Hence the goal is to increase the awareness of consumers own electricity consumption such that they can deduct connections between their consumption and appliances and thus decrease their consumption if possible. The motivation for energy companies to do this is because Danish energy/electricity companies are obligated to help consumers decrease their consumption [9], hence the main goal of gamification in this setting is to investigate if an approach can be quantified.

From this we might gain a notion of the context. Actions that are meaningful regarding electricity use. In order to increase awareness of the consumers we could investigate game design for persuasive games.

In [10] Bogost is defining the concept procedural rhetoric which is defined as: the practice of using processes persuasively. Here a process is the logic behind performing an action. Bogost uses this term in the context of video games, which present a situation or case, using procedural rhetoric's, to a player and let the player decide how to act. In this setting an action could be to turn an appliance on/off, changing the electricity use. In Bogost own words "persuasive games are video games that mount procedural rhetoric's effectively".

E.4 Game development

Based on the described gamification approach a formulation of a goal for using gamification in relation to electricity consumption is the following:

> We want to increase awareness of the consumers electricity use behaviour (the consumption pattern) triggering beneficial changes in the behaviour, in such

a way that the change should be an active choice based on the knowledge gained by playing a game e.g. turning off lights or other appliances or even do energy renovations. This could be done be introducing an engaging way of interacting with electricity use through a game.

But we should be aware of pitfalls, e.g. choices based on wrong information due to faulty models. An example of this was a Watts user who had a very low forecast (due to data issues) of his consumption, hence Watts indicated a heavy overuse and he thought his heat-pump was broken and ordered a technician to check it out. Nothing was wrong and his consumption was normal.

To describe the setting where the consumers can be engaged a short description of the app Watts is given. There are three main features in Watts:

- (Self-)Monitoring: The main feature in Watts is self-monitoring of the consumption compared to a quarterly forecast.
- Comparison with similar households: A first implementation of comparison between similar household is present in Watts, changes based on the current quarterly estimate of consumption.
- CO2 monitoring: The current CO2 load per kWh produced in Denmark is shown and the estimated load for each hour of the day.

In the above features the interactivity is to change the consumption (consumption pattern) which is an indirect way of interacting, there is no engaging game design hence the current interaction is obfuscated. From the current features in Watts we have an effective monitoring tool, but there is not an engaging way of interacting. In this setting the target audience is potential all Watts users, which is a huge and diverse population.

E.4.1 Inspiration

Tamagotchi. A virtual pet that needs care in order to progress in its life cycle and be happy, otherwise it can get sick and/or die, Tamagotchi is currently having a revival [11] and numerous imitations in many different forms have emerged over the years. This shows that nursing and training an entity have proven to be a solid game concept.

Cookie clicker is an incremental game where cookies are produced by clicking. Cookies are accumulated and can be used to purchase buildings and improvements that increase and automate the cookie production. Various achievements are earned with a hint of dark humour, when certain production goals are meet, either for the cookies or the improvements as an incentive to keep playing [12] and [13]. This game shows that accumulating something and then spending it to accumulate even more can be an engaging experience.

> *Cow Clicker is a Facebook game about Facebook games by the author and game designer Ian Bogost. It was first released in July 2010 as both satire and playable theory of social games circa that era.*
>
> *You get a cow. You can click on it. In six hours, you can click it again. Clicking earns you clicks. You can buy custom premium cows and timer overrides through micropayments. Cow Clicker is Facebook games distilled to their essence. [14]*

Given the unexpected “success” of Cow Clicker [15], we have an indication that a waiting time before continuing the game not necessarily would discourage playing a game.

E.4.2 Development of current game design

In the development process the first ideas where merely extended features to the app itself. The following are three early ideas.

- Build a consumption pattern: how could consumption patterns look like for a simulated household where the user chooses the input (house type, appliances, etc.). This could be compared to the users own consumption.
- Play pretend with what if scenarios (price, heat-pump etc.), how would your electricity bill look like if you played with the price installed solar cells or a heat-pump? This could again be compared to the user’s real consumption/bill.
- An energy Tamagotchi. An entity which need to be sustained by energy but depending on time, CO2 load or something else. The entity could have certain preferences such that a household needs to shift/change the consumption patterns in order to keep the Tamagotchi healthy.

From these ideas a first game concept was developed. Small entities need to be sustained in order to survive and prosper. The entities need to be sustained be the players own electricity consumption, this is done be betting the entities on how the consumption is tomorrow. If the player is correct the entities will prosper otherwise, they will die. By betting on tomorrows consumption, the player can interact with the game by using appliances at a certain time or not, the following day.

Starting the development of the game one major limitation became clear. The data of electricity consumption are usually two days delayed which means that that the interaction with electricity use is separated in time from the game. Furthermore, setting up processes for fetching new data automatically proved a too cumbersome task for a prototype game. Hence the prototype developed uses the latest data observed as input and the actual interaction with the game be using electricity after the entities are placed is lost. With these limitations the game becomes a memory game where the player needs to remember what has happened in the household the previous days in order to sustain the entities.

The following is a description of the prototype game developed called PowerMons. First the key elements of the game are described later the game play and last how the data is used is described.

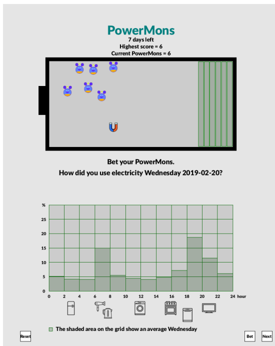
- PowerMon: An entity that needs electricity to be sustained and multiply. It also represents a charge of the battery.
- The battery: The habitat for the PowerMons to frolic when they aren't assigned to the grid for feeding.
- The Grid: The Grid is where the PowerMons are being placed in order to feed and possible multiply/breed. The Grid is split into 12 columns one for every 2 hours of the day. this makes a square grid of 72 slots where one PowerMon can occupy one slot. each slot represents roughly 5% of the daily power use.
- The Magnet Tool: Since the PowerMons are electricity entities, the idea is that an electromagnet is needed to move them. When the mouse-key is pressed in the Battery all the PowerMons there are attracted to the magnet, multiple PowerMons can be picked up.

The game starts with an intro screen (Figure E.1a) introducing the player to the game. It is a turn based game, at the start of the turn the player has to place all PowerMons from the battery into the grid, distributing them as the expected relative electricity consumption (the percentage distribution) will be distributed (see Figures E.1b and E.1c).

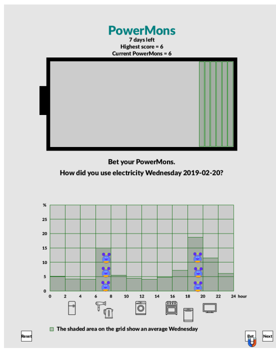
The game starts with six PowerMons so the player can lose a couple in the beginning and still get back into the game. As a help the average of the same weekday's distribution is shown on the Grid (a green shaded area). When a PowerMon is placed in a column it will automatically move to the bottom slot not occupied. When all PowerMons are placed the "bet" button is pressed and the actual consumption distribution is shown (see Figure E.1d).

If there is not enough power in a slot where a PowerMon is placed, it dies. If there is enough power, it is sustained. If there are more than one PowerMon in a column and more than one is sustained, an extra PowerMon is appearing in the battery for every two sustained PowerMons in the column. If there is two

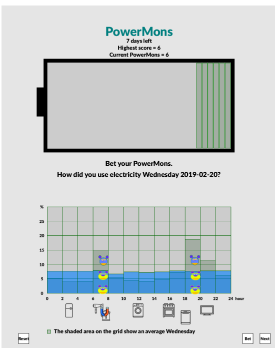
(a) The intro screen briefly describes how the game works.



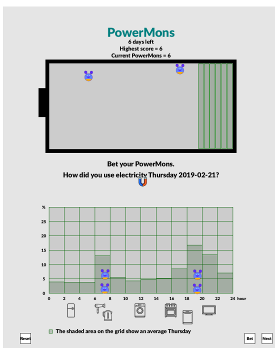
(b) Start of the game. The PowerMons needs to be dragged from the battery down to the grid using the magnet tool.



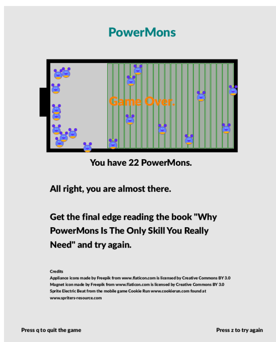
(c) PowerMons are placed in the grid were the player think his consumption was on the day specified above the grid.



(d) The bet button was pressed and the actual consumption was revealed. Two PowerMons were not sustained and will disappear. Since two columns have two PowerMons which are sustained in each of them, two new PowerMons will appear in the battery the next day, one for each column.



(e) The next button was pressed and the game progressed to the next day. Two new PowerMons have appeared in the battery. The sustained PowerMons from the previous day are still placed in the grid and can be moved to new slots.



(f) When the game is over the result is presented to the player with an encouragement to try again if the highest score is not achieved.

Figure E.1: The game play illustrated.

sustained PowerMons one new is appearing. if there are three sustained in one column, two new appear.

To continue to the next day the "Next" button is pressed (see Figure E.1e). In total the game last seven days and the player's performance is measured in the number of PowerMons accumulated during the game. The game is won if the battery is fully charged at the end of the game. The number of PowerMons determine the charge of the battery. A fully charged battery is 30 or more PowerMons. In the end game screen (Figure E.1f) the result is shown.

The consumption shown in the game is the percentage consumption calculated for each column in the grid. Using the percentage distribution, it will always be possible to sustain 18-24 PowerMons no matter how much or little electricity is used. This is also why the winning score is set to thirty. This also increases the difficulty when more PowerMons are earned, since more precision is needed placing them for all of them to be feed. If more than 30 % of consumption is in one column, then the above 30% consumption is lost due to the limit of six slots in one column. This is to limit the possibility that players could change their consumption to only one column which is not the intention of the game. To help the player an average distribution is calculated for each week day the game is played, based on the players own consumption.

We believe that this prototype constitutes a procedural game by the definition given in Sections E.2 and E.3.

E.5 Game testing

The approach for testing with Watts users is based on [4] (Ch. 9), and it was decided to divide the test as follows:

- Introduction
- Warm-up Discussion, questionnaire on how the player is using Watts
- Play session
- Discussion of Game

In the warm-up discussion a questionnaire is given to the testers about their use of the app Watts. This is to get an indication of their knowledge of their own consumption. during the play session the testers are not helped unless they are completely stuck in the game. The testers voice and the screen with the game is recorded during the play through and the tester is encouraged to speak out-loud their experience. In the discussion of the game both questions about the game and game play are asked, furthermore questions on their electrical appliances impact on their consumption is asked to get an indication if a relation between the game and consumption is noticed by the tester. The questionnaires are summarised in Tables E.1 and E.2

It is desired to test if the game was playable and if the relation between actions in the household and electricity consumption can be illuminated through the game experience.

Prior to testing the game with actual users, self-testing was used in the iteration process until a prototype was finished enough for testing. This step was reached when there was a smooth game-play and no major flaws in the controls. The self-testing process is very subjective and things that seems obvious for the game developer can be very counter intuitive for actual users.

For play testing when not self-testing the approach have to be more organized. Since we are both interested in the playability of the game but also if the testers can relate the game to their consumption, we need to be careful in the chosen questions not too lead the testers in to the connection between the game and their consumption.

E.5.1 Testing with confidants

The first testers where 2 colleagues. the set-up was to throw the players into the game without any introduction beside the intro screen. With very little guidance they quickly learned the game mechanics and they seemed engaged trying to remember what they had been doing the last week. One observation on their approach to starting the game was that the intro screen was quickly skipped. So, the ease which the players got engaged in the game could also be due to prior knowledge of the concept of the game. From this preliminary test the players talked about their consumption hence it is suggested to record the play testers while they are playing to hear if they are talking about their consumption and later ask them if they could relate the consumption distribution to their actions with appliances in the home.

The second test with a confidant was a family member who is a Watts user. This test was conducted to see if the play-test manuscript should be altered before testing with the target audience. It was not found to be altered and the result of the test with confidant is shown together with the result of the test with target audience. It was not found necessary to change the test manuscript.

E.5.2 Testing with Watts users

The game was tested using 4 testers, 2 females and 2 males, all employed at SEAS-NVE, working in other departments than the author, and having the app Watts installed on their phones. All the testers have previously been testing other features in the app Watts. This is not representative for the whole population of Watts users, but indications on playability and relation to own consumption is expected to surface with this test.

In Table E.1 a summary of the testers prior app use is presented, this gives an indication of their prior knowledge of daily electricity use. From this table it is observed that all the testers seem to have a good knowledge of their consumption.

Table E.1: Summary of questions prior to play session

	Tester			
Question	Female 1	Male 1	Female 2	Male 2
When did you last use the app Watts?	Five days ago, to check if the consumption had risen during last weekend due to a dryer.	Yesterday, showing to a second party.	One month ago.	Three days ago.
How often do you use the app Watts? (Every day, week, month, quarter or rarer)	Sometimes three times a week, else every other week	Once a week, to check the budget deviation	Once each quarter, usually looking at daily consumption, sometimes hourly.	Every second week, also in work related settings
On a scale from 1 (low) to 5 (high), how would you describe your knowledge of your daily electricity consumption pattern?	4, the tester also looks at hourly consumption in the app.	3, based on the knowledge of presence in the household. Does not look into hourly consumption.	3-4, have a clear idea about when in the day the most consumption occurs.	5, have knowledge of hourly consumption and the base load.
Mention 4 appliances which have great impact on your daily electricity consumption?	Stove, dryer, X-BOX and dishwasher.	Stove, dishwasher washer and dryer.	Stove, computer, washer and fridge/freezer.	Stove, wood pellet stove (heating). The tester emphasizes that the appliances must be used every day with a significant impact hence only two can be mentioned

E.5.2.1 Summaries of play session

The play session was one play through of the seven most recent days for which data was available. The sessions took between 7 and 10 minutes.

Female 1 The tester began the session with stating that she only had played the computer games Candy Crush and Hay Day prior to this game test.

In the beginning she expressed that the icons underneath the grid maybe indicated actual appliance use. She also tried to place the PowerMons at specific times of the day. She did not seem to realise that the time slots could only hold one PowerMon in each row. She did not move already placed PowerMons when beginning with a new day. During the play through the tester expressed some thoughts on activities that had occurred later than expected. In general, the game controls seemed to be difficult. She ended the game with a score of 9.

Male 1 The tester started by reading the intro screen thoroughly. During the play through the tester realised that the difficulty increased with the number of PowerMons accumulated and was reluctant to place them, but quickly figured out that it was needed to continue. In the middle of the play-through the consumption was distributed differently than expected and he lost many PowerMons, but quickly realised why. His wife had been at home all day.

In general, there was a lot of thoughts on where the family had been but not on specific appliances. He ended with a score of 25 and seemed pleased with the result.

Female 2 The tester seemed to read through the intro screen thoroughly, started playing but did not say anything, and I asked her why she placed the PowerMons as she did. It seemed that she thought she were doing something wrong, so I told her that I just wanted her to speak out what she was thinking. She continued and did not say anything. half way in the play through she asked what is the blue area showing up when I press "next" and what is the green area showing? I told her that the green area represented the average distribution of her consumption on Thursdays, since the current day in the game was a Thursday and that the blue was the actual distribution. She then realised that she needed to figure out what she did last Thursday. In the last day of the play through she realised that the days were changing every time she pressed next, and got very concentrated placing the PowerMons, ending the game with a perfect score of 30. In the discussion of the game she was puzzled that the PowerMons looked like mice.

Male 2 The tester started by reading the intro screen and started playing the game. Did not say much in the first half of the play through but in the middle stated that he didn't understand what was going on. "It does not make sense.

It is base load in the middle of the day it is not power monsters?". Continuing that he did not understand the green columns or when the sprites where flashing if it was good or bad. He finished the game placing the PowerMons at what seemed to be at random, ending with a score of 20.

E.5.2.2 Discussion of game experience

In Table E.2 the questions and answers of the follow up discussion after the play session is presented.

Table E.2: Summary of questions after the play session

	Tester			
Question	Female 1	Male 1	Female 2	Male 2
What is your overall impression of the game? Why?	The idea is good, but difficult to manoeuvre.	Funny way to look at consumption	Confusing that you had a magnet to control with. Difficult to figure out the meaning of the game.	Confusing, needed a clear description on how to play the game and how it works. Monster is something big, it does not make sense for me in relation daily electricity use.
Did you learn to play the game quickly or was it difficult? Why?	It was difficult, because it was difficult to move the PowerMons.	Yes, it was easy, but where did the extra monsters come from?	Pretty fast, but did I do it correctly?	Without an introduction it becomes "I don't want to do this".
Which information could have helped you to begin the game?	Better description/explanation of the game controls and that only one PowerMon can be placed in each slot in the grid	It was not clear how the PowerMons multiplied given the number powered in each column.	Explanation on how the number of the mice works.	Clear introduction and explain the results. Show how the game is played using another player
What was the goal of the game?	To place the PowerMons where the consumption had been.	To charge the battery as much as possible, by knowing the consumption more than a week back in time.	To place the mice on the consumption, to become aware when consumption occurs.	Difficult to understand, to draw attention to consumption. I did not notice the battery.
Was there anything that confused you? Why?	The controls where confusing, didn't realise that already placed PowerMons could be moved again.	The controls were confusing, how the movement of the PowerMons worked.	The average consumption was confusing.	When the mousekey was pressed the PowerMons gathered around the mouse pointer.
What did you like most about the game? Why?	That the actual consumption is shown after betting.	How high can you charge the battery, you really have to think to not make mistakes	That you could move several mice at a time	Awareness of the distribution, but it needs a purpose. Can you move some of the consumption for example?
What was the shaded green area in the grid box?	what shaded area? the actual consumption.	An ordinary weekday	My average consumption	It didn't make sense, explanation is needed.
What was the blue area in the grid box?	Blue?	How the day actually was, how much electricity I used.	Uncertain where it was.	*did not ask*
Which appliances have contributed to the greatest spikes in your consumption the latest days?	Probably the Xbox otherwise less consumption was expected in the evening.	Cooking, the stove.	Our Stove	*did not ask*

From the discussion it was difficult for the testers to understand how the game worked. All testers suggested to improve the introduction screen with better explanation of how the game is played. There was a clear indication that the shown distributions were not interpreted as the percentage distribution, in the following discussion three of the testers did not think it would have made a difference if it was the actual consumption. Furthermore, the magnet tool function was difficult to understand and how the PowerMons multiplied. The goal of charging the battery was only clear for one of the four testers.

E.6 Conclusion

The game was very difficult to understand for the testers in the current form, mainly due to an introduction that did not work as intended. One of the four testers seemed to grasp some the game play but there were still things he did not understand.

Based on the game test we can conclude that a clearer introduction of the game is needed. This could be a tutorial where feedback about 'illegal moves' are provided in-game. If the game mechanics are not understood it is difficult to test if the testers can relate the game experience to their own consumption. We cannot conclude if this game could possibly affect consumption behaviour, but we have an indication that the testers could relate their presence in the home to the consumption pattern and in some of the cases betting their PowerMons accordingly, this is based on the observations of the play-through.

Credits

Appliance icons made by Freepik from `www.flaticon.com` is licensed by Creative Commons BY 3.0

Magnet icon made by Freepik from `www.flaticon.com` is licensed by Creative Commons BY 3.0

Sprite Electric Beat from the mobile game Cookie Run `www.cookierun.com` found at `www.spriters-resource.com`

Acknowledgement

I would like to give thanks to Daniel Cermak-Sassenrath from IT University of Copenhagen for the introduction to game development, the many discussions on gamification and guidance throughout the game development process.

This work is partly funded by Innovation Fund Denmark (IFD) under File No. 5189-00045B.

References E

[1] *Watts*. https://watts.seas-nve.dk/. Accessed: 2019-05-28.

[2] D. Johnson et al. “Gamification and serious games within the domain of domestic energy consumption: A systematic review”. In: *Renewable and Sustainable Energy Reviews* 73 (2017), pp. 249–264.

[3] K. Salen and E. Zimmerman. *Rules of play: Game design fundamentals*. MIT press, 2004.

[4] T. Fullerton. *Game design workshop: a playcentric approach to creating innovative games*. AK Peters/CRC Press, 2014.

[5] S. Deterding et al. “From Game Design Elements to Gamefulness: Defining "Gamification"”. In: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*. MindTrek ’11. New York, NY, USA: ACM, 2011, pp. 9–15. ISBN: 978-1-4503-0816-8. DOI: 10.1145/2181037.2181040. URL: http://doi.acm.org/10.1145/2181037.2181040 (visited on 04/18/2017).

[6] A. Duranti and C. Goodwin. *Rethinking context: Language as an interactive phenomenon*. 11. Cambridge University Press, 1992.

[7] P. Dourish. “What we talk about when we talk about context”. en. In: *Personal and Ubiquitous Computing* 8.1 (Feb. 2004), pp. 19–30. ISSN: 1617-4909, 1617-4917. DOI: 10.1007/s00779-003-0253-8. URL: https://link.springer.com/article/10.1007/s00779-003-0253-8 (visited on 06/01/2017).

[8] S. Deterding. *Eudaimonic Design, or: Six Invitations to Rethink Gamification*. SSRN Scholarly Paper ID 2466374. Rochester, NY: Social Science Research Network, July 2014. URL: https://papers.ssrn.com/abstract=2466374 (visited on 04/18/2017).

[9] Energistyrelsen. *Energiselskabers Energispareindsats*. https://ens.dk/ansvarsomraader/energibesparelser/energiselskabers-energispareindsats. Accessed 2017-06-14. May 2016. URL: https://ens.dk/ansvarsomraader/energibesparelser/energiselskabers-energispareindsats (visited on 06/14/2017).

[10] I. Bogost. *Persuasive Games: The Expressive Power of Videogames*. en. MIT Press, June 2007. ISBN: 978-0-262-02614-7.

[11] *The original Tamagotchi*. http://www.bandai.com/tamagotchi/. Accessed: 2019-01-23. 2019.

[12] *Cookie Clicker*. http://orteil.dashnet.org/cookieclicker/. Accessed: 2019-01-24. 2019.

[13] *Cookie Clicker, the internet's most pointlessly addictive game, is also its most subversive.* https://kernelmag.dailydot.com/issue-sections/staff-editorials/15694/cookie-clicker-capitalist-dystopia/. Accessed: 2019-01-24. 2016.

[14] *Cow Clicker.* http://www.cowclicker.com/. Accessed: 2019-01-24. 2017.

[15] *The curse of Cow Clicker: How a cheeky satire became a videogame hit.* https://www.wired.com/2011/12/ff-cowclicker/. Accessed: 2019-01-24. 2012.

Paper F

Does using the app Watts influence electricity consumption? An investigation using linear mixed effect modelling

Authors:
Jon Liisberg, Jan Møller, Peder Bacher

Technical report:
DTU library.

Does using the app Watts influence electricity consumption? An investigation using linear mixed effect modelling

Jon Liisberg[1 2], Jan Møller[1], Peder Bacher[1]

Abstract

Using linear mixed effect models the effect on electricity consumption of using the app Watts is investigated. A significant decreasing effect on electricity consumption was found in relation to the number of times the app is used per month.

F.1 Introduction

In late 2016 the app Watts (Figure F.1) [1] was launched as a digital energy assistant for electricity costumers at SEAS-NVE [2]. Over time several users have given feedback on the app, telling how they have saved electricity, monitoring their consumption based on the prediction presented in the app. The aim of this report is to quantify these savings.

As of late 2017 data on app use have been collected. Every time the app is opened a record of this is saved. Given this collected data on app sessions, it is of interest to investigate if the data can tell a similar story as the users, i.e., if app use (monitoring consumption) have a reducing effect on the electricity consumption.

The focus will be on single family houses, but apartments and summer houses will also be briefly discussed.

[1]DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark
[2]SEAS-NVE, Digital Innovation, Hovedgaden 36, DK-4520 Svinninge, Denmark

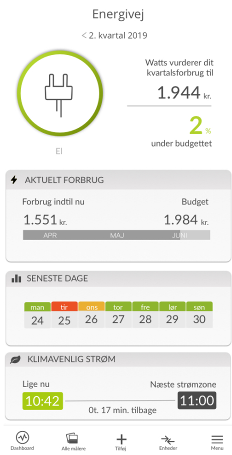


Figure F.1: The Dashboard of the app Watts.

F.2 Data

The data available to investigate the effect of app use is a set of 9000 households with one year of electricity consumption data prior to the installation of the app and at least 3 months data after the installation of the app. For all the households the consumption data was aggregated to monthly observations.

For the reminder of this study the focus will be on single family houses and the following households were removed.

- very low consumption households ($<$ 10 kWh in one month)
- very high consumption households($>$ 3000 kWh in one month)
- unknown house type (880)
- Apartments (757)
- Summer houses (1712)
- Unknown heating type
- Households with solar production

To avoid misclassified summer houses households with very low consumption in one of the observed month is removed. Similar are some enterprises using

Watts and to avoid heavy consumer households with very high consumption is removed,For some of the households the primary heating type was not reported and these have also been removed. Furthermore several households also had production of electricity using solar-arrays which influences the consumption these have also been removed.

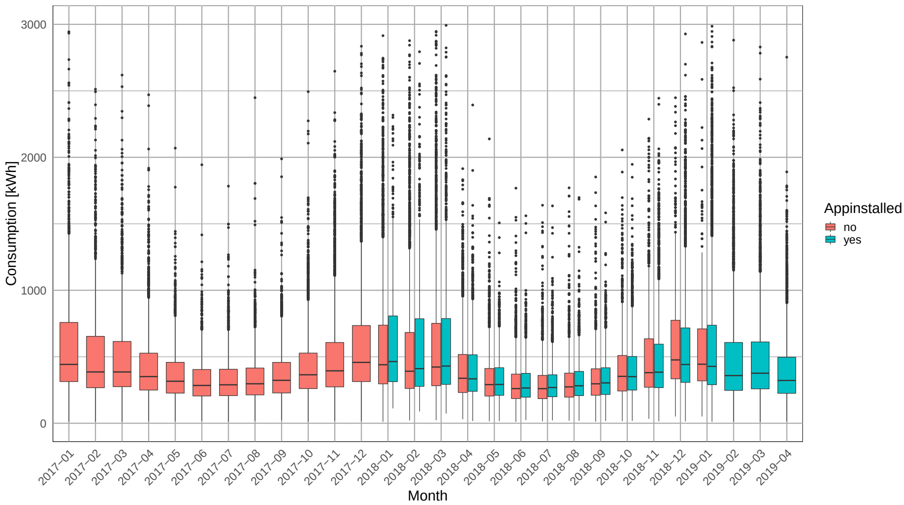


Figure F.2: Boxplots of monthly aggregated consumption given app installation information.

After this 4292 single family households were left with a total of 97568 monthly observations, 55674 prior to the app was installed and 41894 after the app was installed. Box-plots of the consumption are presented in Figure F.2. From this a yearly variation in the consumption is observed, but no clear indication of an effect of having the app installed is observed.

For each month after the app was installed the number of sessions with the app was counted, i.e., every time the app was opened. Prior to app installation the app session was set to zero. Looking into the count of app sessions per month in Figure F.3 a clear effect is not observed. A decrease of the outliers is observed but it could be due to the decrease in the number of observation, in each interval when the app use increases.

No effect of having the app installed or the use of the app can be observed directly, hence further investigation of the data is needed. A clear yearly variation was observed for the consumption, hence looking into the effect of primary heating type is done in Figure F.4. Here it is clear that electrical heating have a huge effect, but also variation between other heating types are observed. Taking the

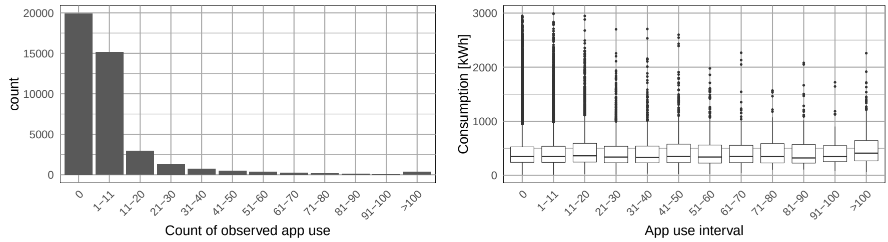


Figure F.3: Left: Count of app use when the app is installed, divided into intervals. Right: Boxplots of monthly aggregated consumption given app use interval.

heating type into account the ambient temperature should also be considered.

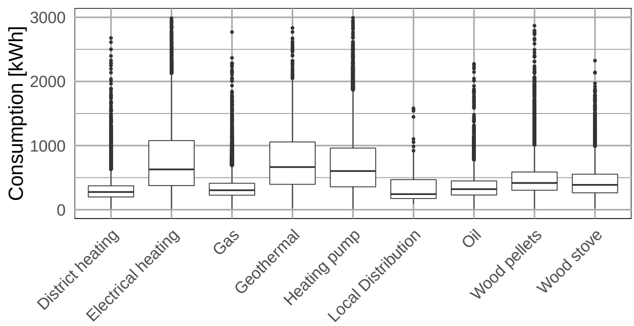


Figure F.4: Boxplots of monthly aggregated consumption given primary heating type.

For temperature input, data from 3 locations on Zealand were averaged for each month resulting in one time-series of temperature. The majority of the households in this study are located on Zealand but some are also spread out in the whole of Denmark. Instead of using the averaged temperature a temperature difference was calculated between a fixed temperature and the averaged ambient temperature as in Equation (F.1),

$$T_{\mathrm{d},t} = \begin{cases} T_{\mathrm{threshold}} - T_{\mathrm{a},t} & \text{if } T_{\mathrm{a},t} < T_{\mathrm{threshold}} \\ 0 & \text{else} \end{cases} . \tag{F.1}$$

This is to have a positive correlation between the temperature input and the consumption dependent on the temperature and also as a possible input describing yearly variation even though it might not be directly correlated. The threshold chosen was 22°C degrees and is presented in Figure F.5. This threshold based on [3](Paper A).

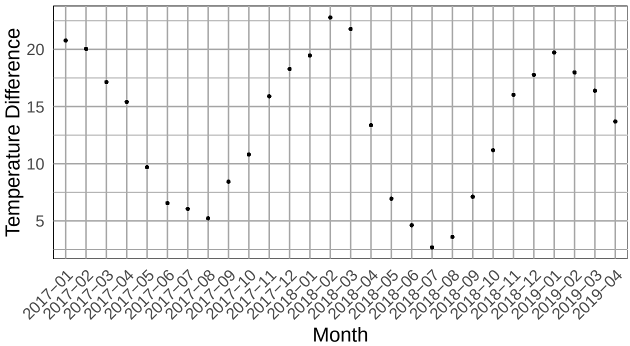


Figure F.5: The temperature difference between fixed temperature of 22°C and the ambient temperature

With a temperature input we can now investigate if an interaction between temperature and heating type should be considered. In Figure F.6 it is clear that there is a increasing effect on the consumption when the temperature difference is high (it is cold), for electrical primary heating type. There is also observed a small increasing effect for heating types Gas, Oil, Wood pellets and Wood stove. This could be due to secondary electrical heating or the yearly variation observed both in temperature and consumption.

Since there is large variation in the consumption between households these will enter the model as random effects.

Based on this exploratory analysis no clear indication of an effect from app use was observed, but clear effect of heating type and temperature was observed. The data chosen as variables in this study are presented in Table F.1. Using primary heating type and the temperature difference as input we might be able to see if there is an effect on the consumption based on the app use.

Knowledge on secondary heating type was also available and similar relation to temperature was observed as for the primary heating type. Since many households had reported the same primary and secondary heating type the secondary is disregarded in this study. Also the size of the house and number of residents were available but many households had not reported it and these possible variables are left out as well.

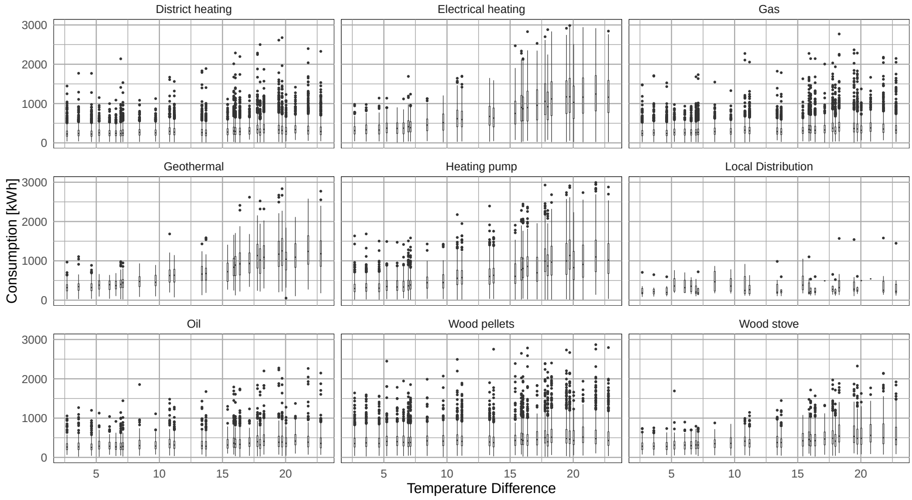


Figure F.6: Boxplots of monthly aggregated consumption given the temperature difference for each heating type.

Table F.1: Variable description

Variable	**description**
Y	The monthly aggregated electricity consumption.
`Appuse`	Count of app sessions within the observed month
`Appusegroup`	Grouping of appuse count in intervals.
`Appinstalled`	Was the app installed at the observation
`PrimaryHeatingType`	Primary heating type
`temp.diff`	The temperature difference between fixed threshold (22°C) and ambient temperature.
`InstallationId`	A meter identifier (household)

F.3 Method

In this study linear mixed effect models are utilised to investigate if the use of the app Watts have an effect on the electricity consumption in the users household. To do this linear mixed effect models will be used, with which the effect of app use can be investigated if it is found significant.

Linear mixed models differ from linear models by introducing random effects in the model description

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}, \tag{F.2}$$

where $\mathbf{y}$ is the vector of observations with mean $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$. $\boldsymbol{\beta}$ is the vector of fixed effect regression coefficients. $\mathbf{u}$ is the vector of random effects ($\mathbf{u} \sim N(0, \Sigma)$). $\boldsymbol{\varepsilon}$ is the vector of residuals ($\boldsymbol{\varepsilon} \sim N(0, \sigma^2)$). $\mathbf{X}$ is the predictor variables. $\mathbf{Z}$ the design matrices for the random effects. For further reading on linear mixed effect models see [4].

The model are implemented in `R`[3] using the `lme4` package [5] for model fitting and the `lmerTest` package [6] for testing the model fits.

F.4 Results

Since the response of the model is assumed normal distributed and the consumption cannot be negative the consumption is log transformed before the model fitting. Given this and based on the exploratory data analysis the following model was formulated.

$$\begin{aligned}
\log Y_i = {} & \alpha(\texttt{PrimaryHeatingType}_i) && \text{(F.3)} \\
& + \beta(\texttt{temp.diff}_i) && \text{(F.4)} \\
& + \gamma(\texttt{PrimaryHeatingType:temp.diff}_i) && \text{(F.5)} \\
& + \delta_1(\texttt{Appusegroup}_i) && \text{(F.6)} \\
& + \delta_2(\texttt{Appinstalled}_i) && \text{(F.7)} \\
& + e(\texttt{InstallationId}_i) + \varepsilon_i && \text{(F.8)}
\end{aligned}$$

Where $i = 1, \ldots, 97568$,

$$e(\texttt{InstallationId}_i) \sim N(0, \sigma_e^2), \varepsilon_i \sim N(0, \sigma^2) \quad \text{(F.9)}$$

all are iid.

`PrimaryHeatingType`, `temp.diff`, `Appusegroup` and `Appinstalled` are fixed effects, and `InstallationId` is the random effects.

After the model fit an ANOVA was conducted on this to see if some of the fixed effects can be removed. The result of this ANOVA is presented in Table F.2, from this it is clear that `Appinstalled` is not significant and should be removed.

Reducing the model and testing it again with an ANOVA (Table F.3), no further reduction of the fixed effects was found needed.

[3] `https://cran.r-project.org/`

Table F.2: ANOVA of proposed model

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
`PrimaryHeatingType`	32.25	4.03	8.00	4924.29	55.93	< 0.0001
`temp.diff`	597.32	597.32	1.00	93278.41	8286.76	< 0.0001
`Appinstalled`	0.07	0.07	1.00	93745.36	0.94	0.3330
`Appusegroup`	13.72	1.25	11.00	93598.66	17.30	< 0.0001
`PrimaryHeatingType:` `temp.diff`	1736.52	217.06	8.00	93268.23	3011.41	< 0.0001

Table F.3: ANOVA of reduced model

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
`PrimaryHeatingType`	32.26	4.03	8.00	4924.42	55.95	< 0.0001
`temp.diff`	597.25	597.25	1.00	93279.46	8285.86	< 0.0001
`Appusegroup`	16.21	1.47	11.00	93571.89	20.45	< 0.0001
`PrimaryHeatingType:` `temp.diff`	1736.69	217.09	8.00	93269.27	3011.71	< 0.0001

To investigate if the random effect of `InstallationsId` is significant or that a linear regression model is adequate a test of the random effect is carried out using the `step` function from the `lmerTest` package. From this test (Table F.4), the random effect of the household was found significant and are kept in the model.

Table F.4: The resulting table from `step` function testing if random effect are significant.

	Eliminated	npar	logLik	AIC	...	Df	Pr(>Chisq)
<none>		31.00	-19165.39	38392.78			
(1 | `InstallationId`)	0.00	30.00	-75090.21	150240.43	...	1	< 0.0001

Since the `Appusegroup` was found significant a model with `Appuse` was fitted reducing the number of parameters in the model. The same model structure was found as for the `Appusegroup` model, with a slight decrease in AIC and BIC (Table F.5), hence `Appuse` is the preferred model.

F.4.1 Residual analysis

To validate the model assumption that the residuals are normal distributed, the plots in Figure F.7 are investigated. The majority of the residuals seems to be normal distributed but from the qq-plot there are observed heavy tales. These tales could be explained by other effect like size of the house and the number of residents but for now these data are not available with high enough quality. Also individual temperature thresholds for each household and more

Table F.5: Information criteria comparison between `Appusegroup` and `Appuse` models

	`Appusegroup`	`Appuse`
AIC	38392.78	38331.93
BIC	38686.91	38531.19

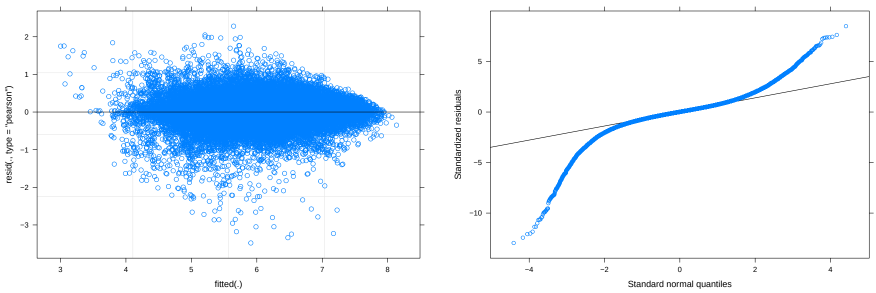


Figure F.7: The left panel shows a plot of the residuals. The right plot shows a qq plot of the residuals

locally observed temperature might improve the fit. Furthermore, since we have several observation over time for each household the residuals might not be independent, this could be investigated by analysing the acf for each household, but as an overall conclusion in relation to the investigation of an effect of app use, there is no strong evidence against this assumption and the model is found adequate. Similar results for the linear `Appuse` model was found.

F.4.2 The effect of heating type and temperature

To show the parameter estimates of the heating effect based on the heating type and temperature, these are presented as the multiplicative effect in Figure F.8. The multiplicative effect is calculated as follow:

$$heateffect_i = \exp(\alpha_i + \beta_i \cdot T_{\text{diff}}), \tag{F.10}$$

where α_i is the intercept parameter, β_i is the slope parameter for heating type i and T_{diff} is the temperature difference for the given month. District heating is contained in the intercept hence the effect of this is 1. In the plot the temperature difference is is reverted back to outdoor temperature, hence the intercept is at 22°C.

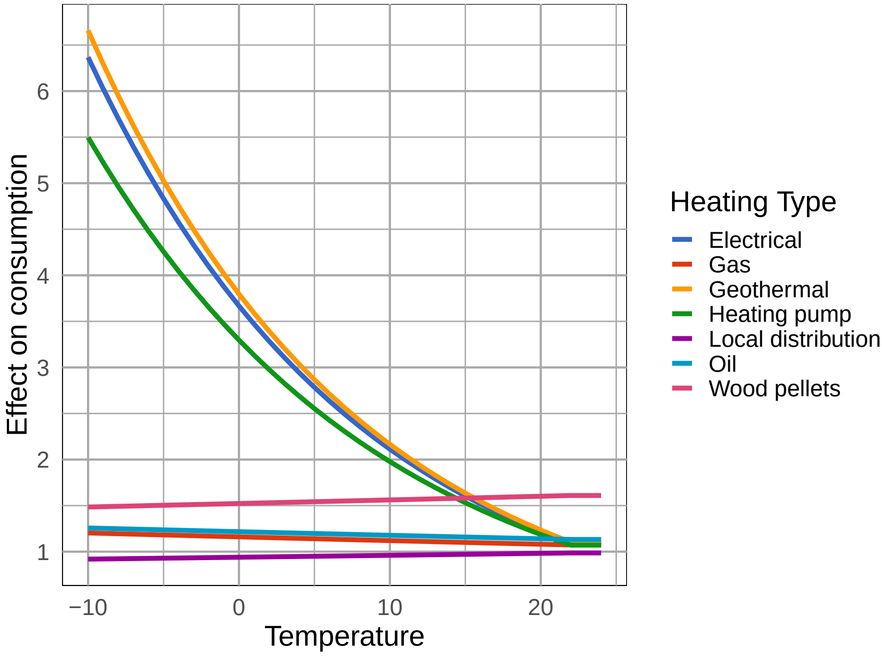


Figure F.8: The multiplicative effect of heating type in single family houses.

F.4.3 App use effect

Since the model is fitted on the log-transformed consumption, we can extract the multiplicative effect of the app use, by using the exponential function on the parameter estimates.

$$Appuse = \exp(\beta \cdot N), \qquad \text{(F.11)}$$

where β is the slope parameter for `Appuse` and N is the number of app use per month. These multiplicative effect are presented in Figure F.9. We have found a significant relation between the number of app sessions per month and the consumption. It should be noted that this effect is not for the individual user but an average effect observed for the whole population. From the plot we can see that if the users interacted with Watts 41-50 times in a month the average effect for this group of users was to reduce the consumption to 95 % of the consumption compared to users not interacting with watts.

The described model was also tested for summerhouses and apartments. For the summer houses an increasing effect of the app use was found (Figure F.11),

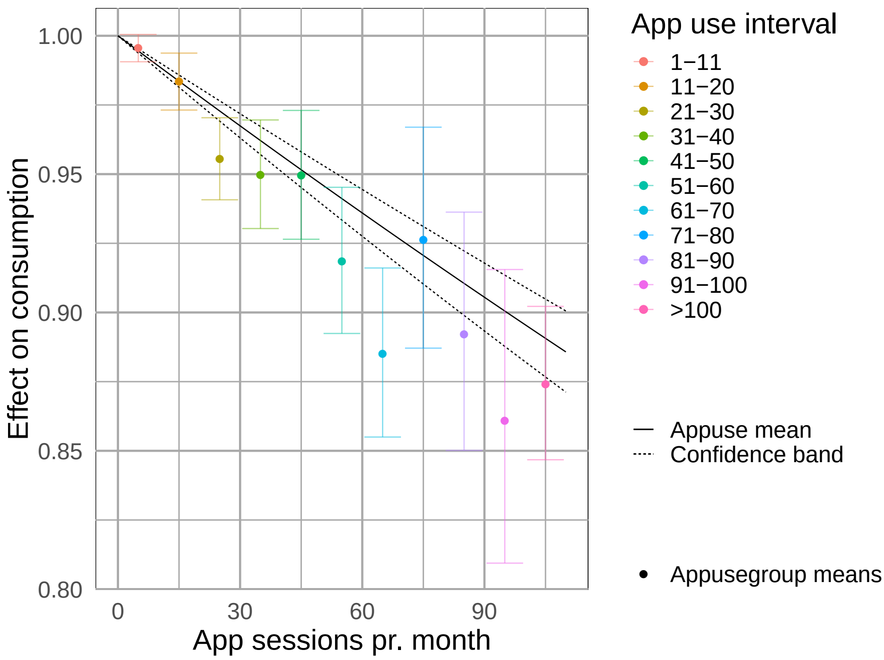


Figure F.9: The multiplicative effect of `Appuse` and `Appusegroup` in single family houses, with 95% confidence intervals

this might be due to an increase in the app use while using the summer house, or just after the summerhouse has been used hence the consumption is much higher than an empty summer house. For apartments (Figure F.10) we did not find a significant effect of app use on the consumption, this could be due to the smaller sample or the limited possibility to lower consumption due to the already low consumption observed for apartments and that a reduction would be low in comparison to the effort put into it.

F.5 Discussion/Conclusion

A clear decreasing effect on the consumption was found for increasing app use. Hence we can conclude that active user have lower consumption than non- or less-active users. We can conclude that the active user on average reduced their consumption, but we cannot conclude an effect for single users only the whole population. It is unclear if the app have inspired users to monitor and reduce their consumption or if the active users where already active in saving electricity and utilize the app as a tool for this. We have found that the the data tells a similar story as the users feedback when considering single family houses.

Furthermore, the random effect could be used as feedback in comparison to the rest of the population since it gives a relative measure of how close the household is to the mean consumption.

Acknowledgement

This work is partly funded by Innovation Fund Denmark (IFD) under File No. 5189-00045B.

References F

[1] *Watts*. https://watts.seas-nve.dk/. Accessed: 2019-05-28.

[2] *SEAS-NVE Om koncernen*. https://www.seas-nve.dk/koncernen/om/koncernen. Accessed: 2019-05-28.

[3] J. Liisberg, J. Møller, and P. Bacher. "Models for long-term baseline prediction of daily electricity consumption in individual households". In: (). submitted for consideration to Applied Energy.

[4] D. Bates et al. "Fitting linear mixed-effects models using lme4". In: *arXiv preprint arXiv:1406.5823* (2014).

[5] D. Bates et al. "Package 'lme4'". In: ().

[6] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen. "lmerTest package: tests in linear mixed effects models". In: *Journal of Statistical Software* 82.13 (2017).

F.6 Apartment and summer house results

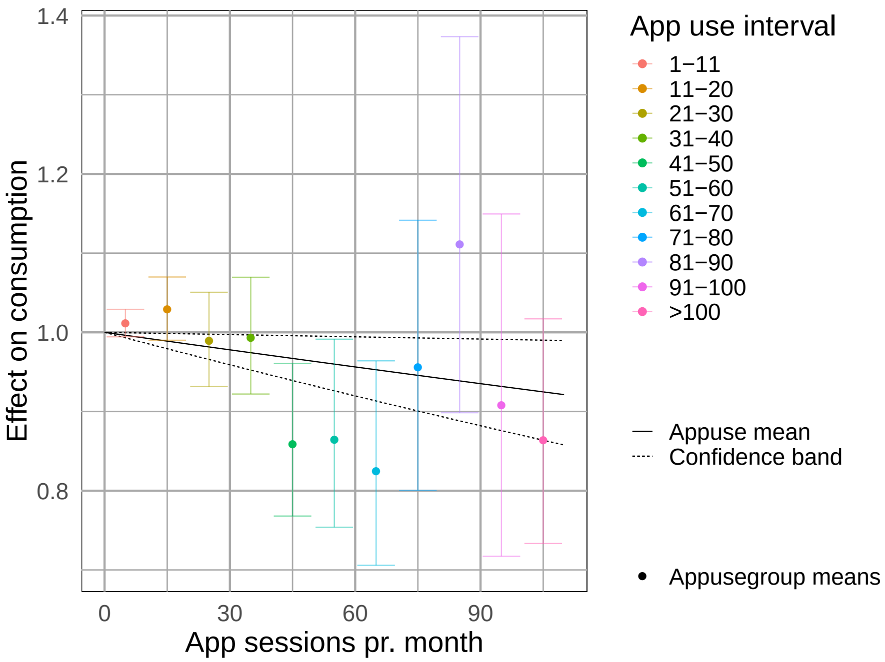


Figure F.10: The non-significant multiplicative effect of `Appuse` and `Appusegroup` in apartments, with 95% confidence intervals

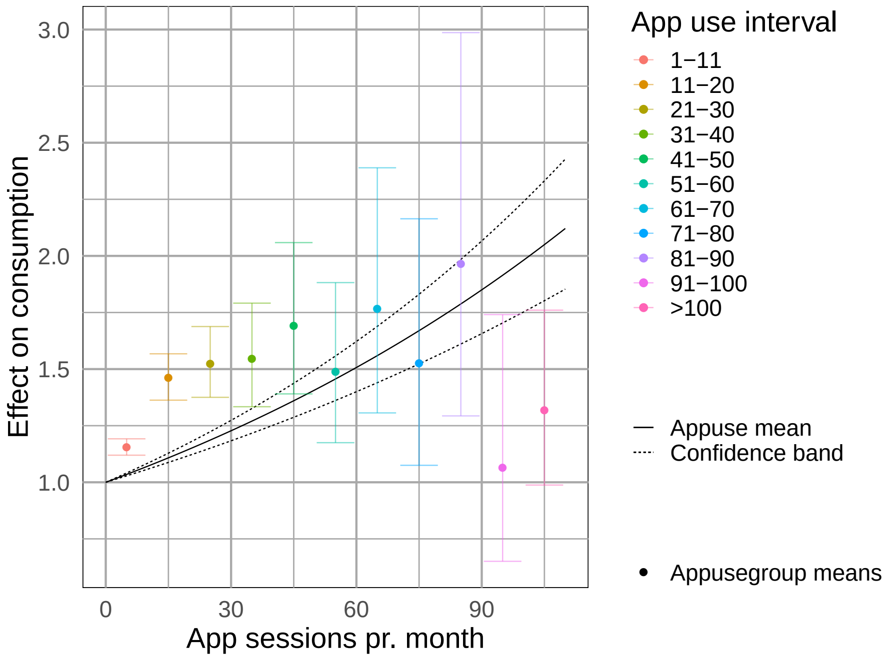


Figure F.11: The significant multiplicative effect of `Appuse` and `Appusegroup` in summer houses, with 95% confidence intervals