



Hidden Markov Models for indirect classification of occupant behaviour

Liisberg, Jon Anders Reichert; Møller, Jan Kloppenborg; Bloem, H.; Cipriano, J.; Mor, G.; Madsen, Henrik

Published in:
Sustainable Cities and Society

Link to article, DOI:
[10.1016/j.scs.2016.07.001](https://doi.org/10.1016/j.scs.2016.07.001)

Publication date:
2016

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Liisberg, J. A. R., Møller, J. K., Bloem, H., Cipriano, J., Mor, G., & Madsen, H. (2016). Hidden Markov Models for indirect classification of occupant behaviour. *Sustainable Cities and Society*, 27, 83-98. DOI: 10.1016/j.scs.2016.07.001

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Hidden Markov Models for indirect classification of occupant behaviour

J. Liisberg^a, J. K. Møller^a, H. Bloem^c, J. Cipriano^b, G. Mor^b, H. Madsen^a

^a*DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark*

^b*CIMNE-UPC, Terrassa, Spain*

^c*European Commission, JRC – IET, RE-Unit, TP 450, I-21020 Ispra, Italy*

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

Keywords: Occupant behaviour, Indirect classification, Hidden Markov Models

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 oversimplified. 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. $t \in \{1, \dots, T\}$
\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
$\mathbf{A}, \mathbf{\Gamma}$	matrices
$\mathbf{a}, \boldsymbol{\theta}$	row vectors
\mathbf{a}'	a column vector

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.

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

$$\begin{aligned} &\text{component distribution } \delta_1, \dots, \delta_m \\ &\text{probability or density functions } p_1(x), \dots, p_m(x) \end{aligned}$$

The component is specified by the discrete random variable C which performs the mixing where $Pr(C = i) = \delta_i$ for $i \in \{1, \dots, m\}$ 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:

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

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.

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

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:

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

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

$$\mathbf{\Gamma} = \begin{pmatrix} \gamma_{11} & \dots & \gamma_{1m} \\ \vdots & \ddots & \vdots \\ \gamma_{m1} & \dots & \gamma_{mm} \end{pmatrix}. \quad (4)$$

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

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

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

Hence, the dynamics is described by the unobserved parameter process $\{C_t : t = 1, 2, \dots\}$, which describes the evolution of the states in time. The observations are described by the state-dependent process $\{X_t : t = 1, 2, \dots\}$ 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 1.

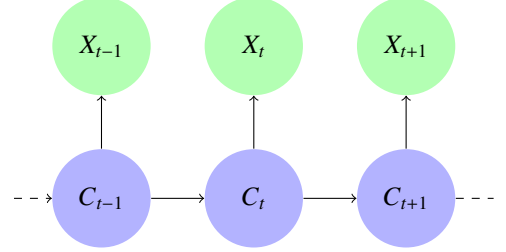


Figure 1: Directed graph of a Hidden Markov Model.

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 $(X_1, \dots, X_{t-1}, X_{t+1}, \dots, X_T)$, then the conditional distributions of X_t given all other observations are given by:

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

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

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

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

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:

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

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

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 $\mathbf{x}^{(T)} = (x_1, \dots, x_T)$ is the observations. The maximum likelihood parameter estimates are found as:

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

Using the Markov property, this reduces to

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

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

$$L(\theta, \mathbf{x}^{(T)}) = \prod_{t=1}^T Pr(X_t = x_t | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}) \quad (12)$$

$$= \prod_{t=1}^T Pr(X_t = x_t | C_t = i) Pr(C_t | \mathbf{X}^{(t-1)} = \mathbf{x}^{(t-1)}) \quad (13)$$

$$= \prod_{t=1}^T \sum_{i=1}^m \delta_i \cdot Pr(X_t = x_t | C_t = i). \quad (14)$$

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

$$Pr(X_t = x_t | C_t = i) = \int_{a_t}^{b_t} p_i(x, \theta_i) dx, \quad (15)$$

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

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

2.2.1. Estimation routine

The estimation is typically carried out by direct optimization of the likelihood $L(\theta, \mathbf{x}^{(T)})$. In \mathbb{R}^1 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 Γ , 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 $i = 1, \dots, m$, 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 Γ , 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:

$$\mathbf{T} = \begin{pmatrix} - & \tau_{12} & \tau_{13} \\ \tau_{21} & - & \tau_{23} \\ \tau_{31} & \tau_{32} & - \end{pmatrix}, \quad (17)$$

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

$$Q_{ij} = \begin{cases} g(\tau_{ij}) & \text{for } i \neq j \\ 1 & \text{for } i = j \end{cases}, \quad (18)$$

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

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 Γ, \mathbf{k} and θ the calculation of the maximum-likelihood parameters can be calculated in two steps:

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

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

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

$${}_t\Gamma = \begin{pmatrix} {}_t\gamma_{11} & \cdots & {}_t\gamma_{1m} \\ \vdots & \ddots & \vdots \\ {}_t\gamma_{m1} & \cdots & {}_t\gamma_{mm} \end{pmatrix}. \quad (19)$$

If we consider the same structure in the transformation as for (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

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

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

$${}_t\tau_{ii} = \beta_i \mathbf{y}_t', \quad (21)$$

this is the part where the seasonality is incorporated. \mathbf{y}_t is a vector of p covariates modelling the seasonality, and β_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

$${}_tQ_{ij} = g({}_t\tau_{ij}), \quad (22)$$

and set ${}_t\gamma_{ij} = {}_tQ_{ij} / \sum_{k=1}^m {}_tQ_{ik}$, 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

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

¹<https://www.r-project.org/>

With this approach of modelling the seasonality in Γ , 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:

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

and then calculating ${}_t\Gamma$ for $t \in \{1, 2, \dots, r\}$ with (23). Additional sine-cosine pairs can be included to model more complex seasonal patterns, if necessary. Using sine-cosine pairs, as in (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].

2.2.3. Covariates in the state-dependent distribution

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

$$\log {}_t\mu_i = \alpha_i y_i', \quad (25)$$

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

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

2.3. Model selection

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.

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 (cdf) (Figure 2). For model S the CRPS is given as

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

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.

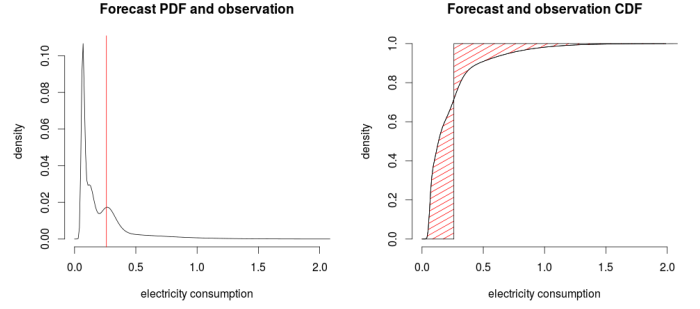


Figure 2: Example of the cdfs used to calculate the CRPS together with pdf and observation.

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.

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

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

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

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

and

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

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.

3. Results

3.1. Data overview

The data are collected in and nearby an apartment building in Catalonia, Spain (Figures A.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.

3.1.1. Weather data

To investigate whether the weather data (Table 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 1: Description of weather data.

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

3.1.2. Smart metering data

Due to a large number of zero observations, the smart metering data (Table 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.

Table 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 liters

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

Table 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 A.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 A.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.

3.2. Homogeneous Hidden Markov Models

The models used for the electricity consumption are first-order Hidden Markov Models (equations 5 and 6), and the structure of this model is illustrated in Figure 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 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 $i = \{1, \dots, m\}$, 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:

$$E[X] = k\theta \quad (30)$$

$$Var[X] = k\theta^2 \quad (31)$$

These are used to interpret and classify the states.

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 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 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.

Table 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
Poisson	2	56	30588	61288	61682
	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

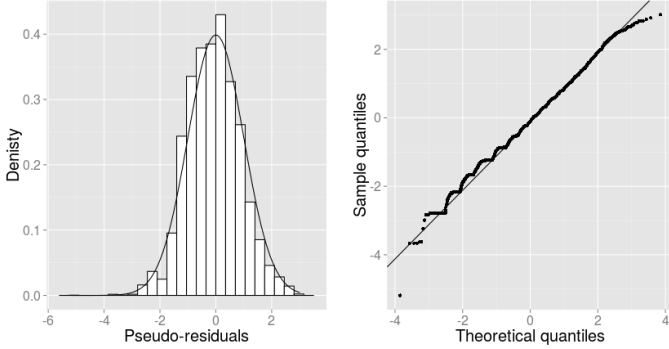


Figure 3: Histogram and qq-plot of the ordinary pseudo-residuals for the three-state HMM, Apartment 2.

To assess the goodness of fit of the three-state model, we look at the ordinary pseudo-residuals (Figure 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 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.

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 (9).

From the average probability profile with respect to time of day, a clear variation/dependence is observed (Figure 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.

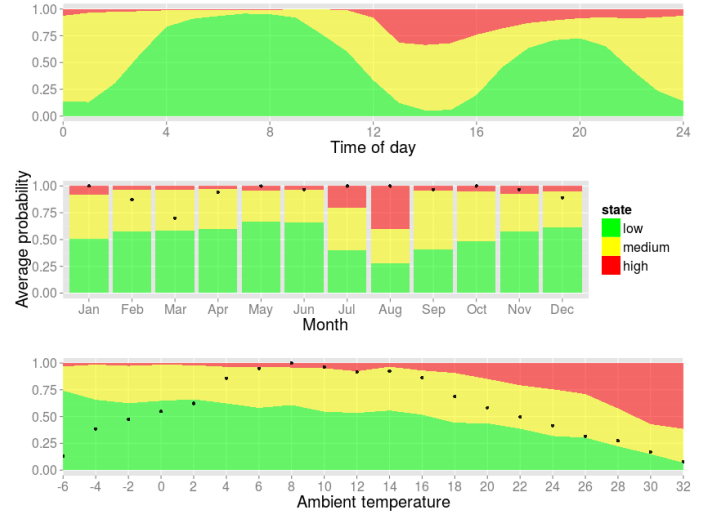


Figure 4: Average probability profiles of being in given state, dependent on time of day, time of year and ambient temperature, Apartment 2. The black dots denote the relative amount of observations for each month or temperature interval.

Given the average probability profile for time of day and the knowledge from the occupant survey (Table 7), the "low

Table 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

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

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 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.

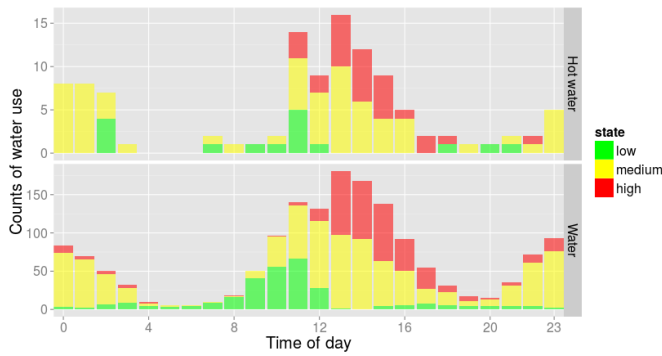


Figure 5: Counts of water and hot-water use given state and time of day, Apartment 2.

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

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 7. The choice of model for each apartment is based on the AIC/BIC values and plots of the state-dependent distributions (Figure A.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.

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 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 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							-

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 6). The four patterns have been classified as shown in Table 8.

For the four apartments with the pattern “mostly at home”, the prior knowledge from the occupant survey (Table 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 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

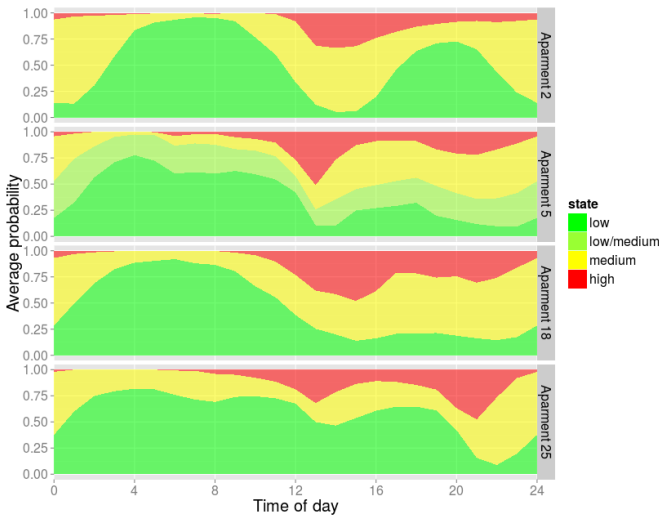


Figure 6: Observed distinct profiles. The low/medium state of Apartment 5 is considered low based on the mean value of the state-dependent distribution.

Table 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 7 and 8).

For the parameter k , huge variation is observed for the low and medium consumption states (Figure 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.

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 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.

3.3. Time-inhomogeneous Hidden Markov Models

In the time-inhomogeneous models, the time dependence is modelled in the transition probability matrix. Estimated param-

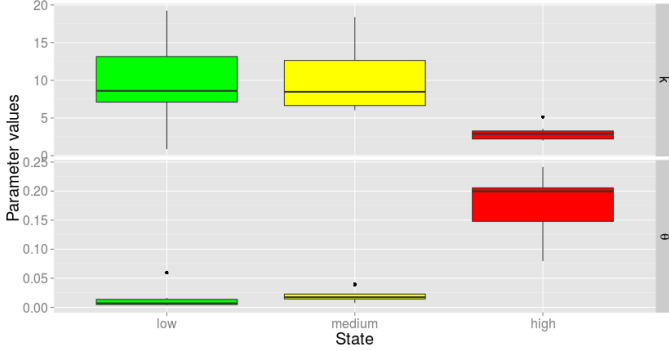


Figure 7: Box-plots of the parameters estimates for the state-dependent distributions, note the difference in the scales.

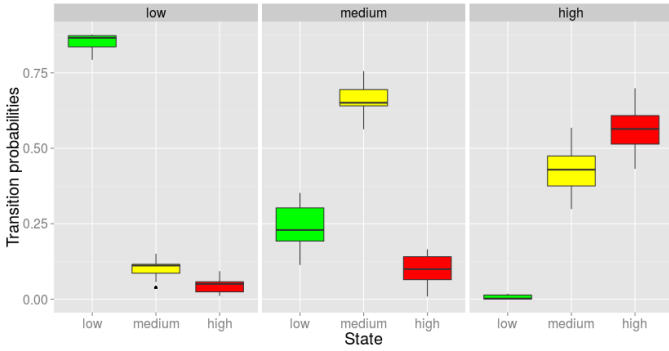


Figure 8: Box-plots of the transition probabilities for each state.

eters are the shape k_i , and the scale θ_i for the state-dependent distributions for $i = \{1, \dots, m\}$, 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 (23), hence the parameters are τ_{ij} for $i, j = \{1, \dots, m\}$ and $i \neq j$, and $\tau_{ij} = \beta_i \cdot y'_t$ for $i = j$ where β_i is a vector of Fourier coefficients for state i and y_t is a vector of sin- cosine pairs to model the daily variation in the states. Hence the number of β parameters depends on the length of y_t and the number of states.

The Fourier series are

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

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 $m(m + 2p + 1)$

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.

Table 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

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

When checking the goodness of fit, by assessing the normal pseudo-residuals (Figure 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.

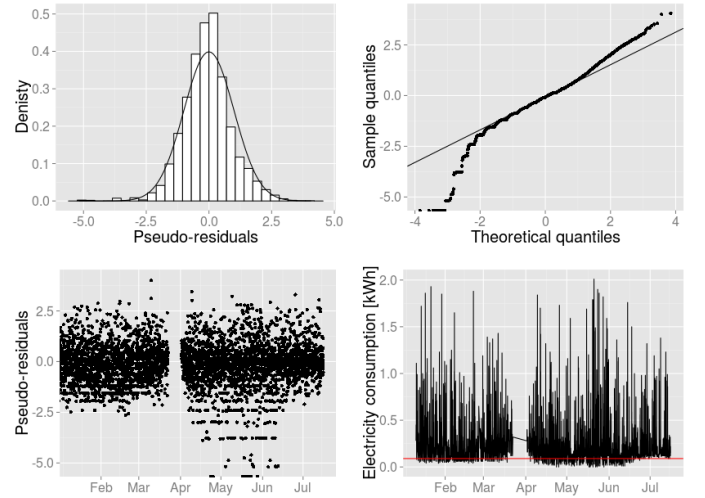


Figure 9: Histogram and qq-plot of the ordinary pseudo-residuals for the three-state inhomogeneous HMM of Apartment 18, also plotted as a function of time. The electricity consumption of the same period of time is shown with the mean of the low consumption state.

To analyse why there is a heavy lower tail, a plot of the ordinary pseudo-residual plotted as function of time is investigated (Figure 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

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

State	k	θ	$\beta_{i1\sin}$	$\beta_{i1\cos}$	$\beta_{i2\sin}$	$\beta_{i2\cos}$	τ_{i1}	τ_{i2}	τ_{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	-0.569	-1.18	-	-1.39
3	2.85	0.21	-0.095	-5.203	2.672	0.202	-0.46	2.18	-

that some kind of filtering or time-adaptive method should enter the model. The parameter estimates are shown in Table 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 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 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.

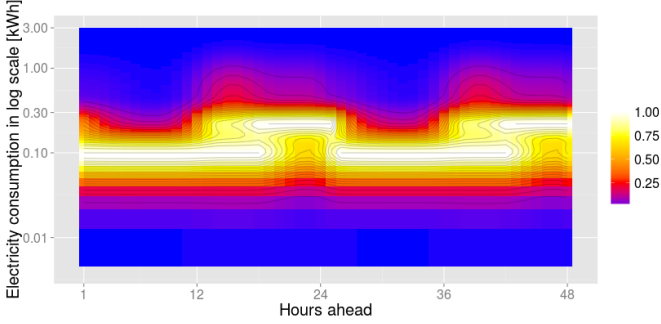


Figure 10: Contour plot of forecasting distributions 48 hours ahead of the data used to fit the model. The scale is relative to the largest probability in each horizon, Apartment 18.

To check whether the inhomogeneous model is a better fit than the homogeneous model, we first look at the AIC and BIC values (Table 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 $h = \{1, 2, \dots, 48\}$, 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 the time-inhomogeneous model and results are then compared (Figure 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.

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

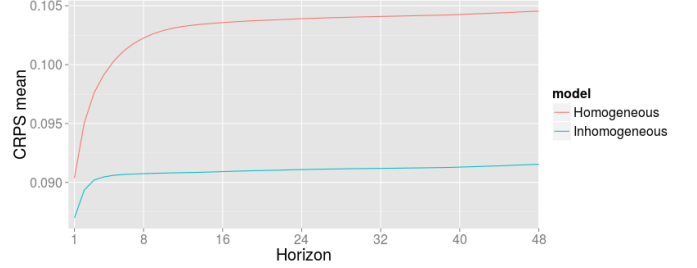


Figure 11: Comparison of the CRPS between homogeneous and time-inhomogeneous models, Apartment 18..

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 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 6), the model-generated daily probability profiles are less accurate than for the apartments where 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.

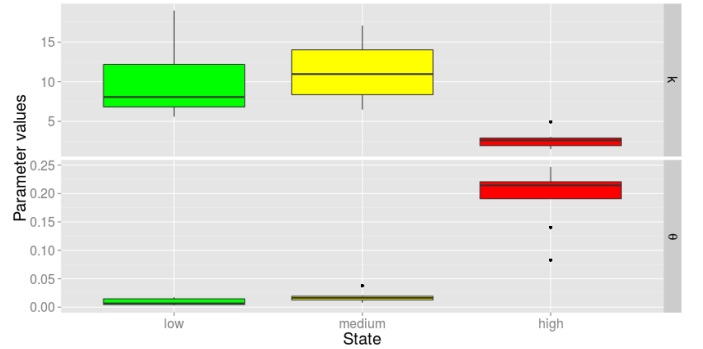


Figure 12: Box-plots of state-dependent parameters for the inhomogeneous three-state models.

Investigating for common parameters, similar observations

are seen for the state-dependent distribution parameter k and θ (Figure 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.

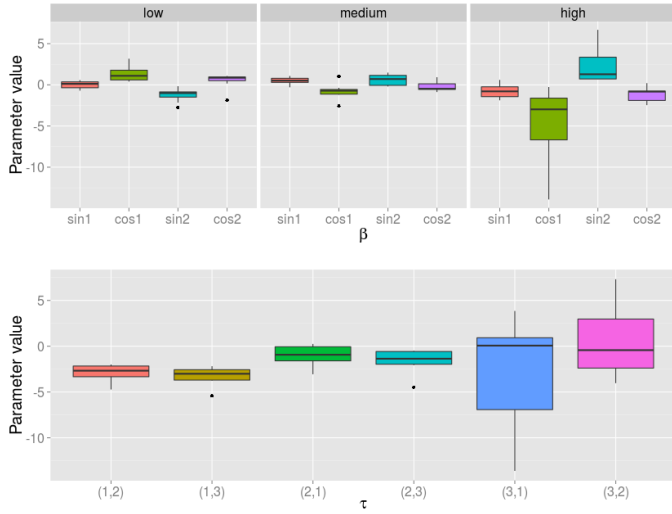


Figure 13: Box-plots of transition probability parameters for the inhomogeneous three-state models.

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.

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 (25), where only the temperature is used as covariate. Hence for each state dependent on temperature, the means are given by

$$\log_t \mu_i = \alpha_{i1} + \alpha_{i2} y_t \quad (33)$$

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 $m(m + 2p + 1) + q$

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 14) with significantly higher mean values in the summer period than the winter period.

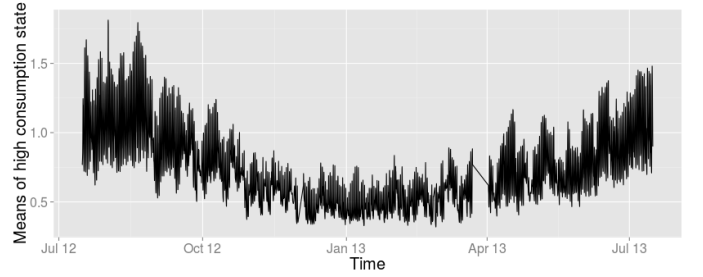


Figure 14: Mean values of the high consumption state over one year.

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 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 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 15), we also see an improvement in the forecasting performance for the time- and temperature-dependent model.

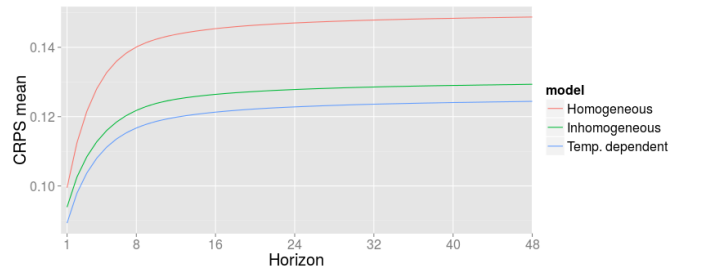


Figure 15: CRPS comparison between the homogeneous model, time-inhomogeneous and the time-inhomogeneous temperature-dependent models, Apartment 2.

When the forecasting distributions are compared (Figure 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 were 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

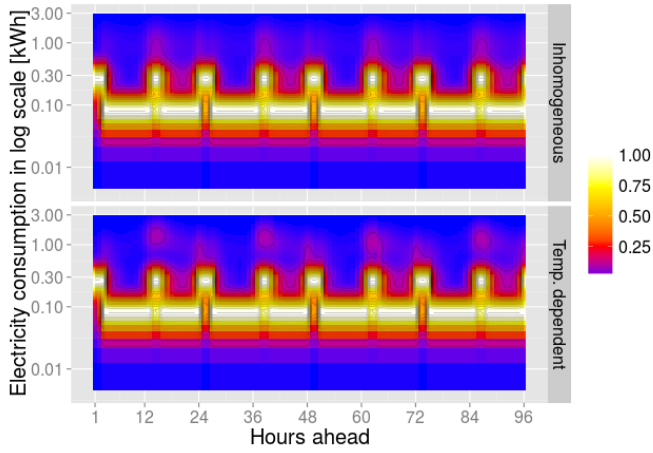


Figure 16: Contour plot of forecasting distributions for the inhomogeneous and temperature-dependent model, Apartment 2. The scale is relative to the largest probability in each horizon. The forecast is in the summer period.

the high consumption state is very low during most of the day, except for the hours in the middle of the day.

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 develop-

ing 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).

Appendix A.

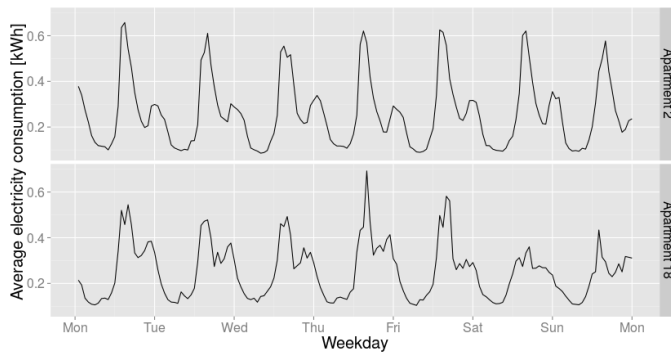


Figure A.17: Weekly averages of the electricity consumption, one year of data was used.

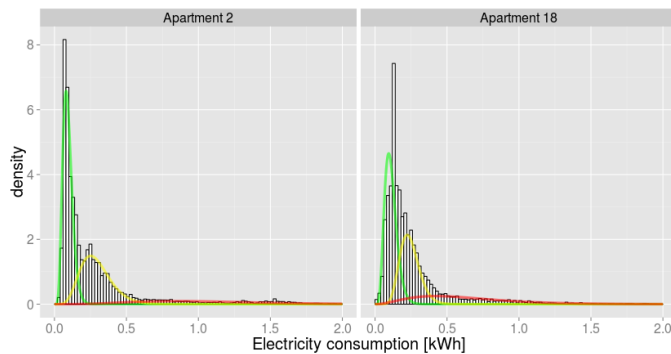


Figure A.18: Histograms of the electricity consumption, one year of data was used. Observations between 2-3 kWh are not shown.

References

- [1] P. D. Andersen, A. Iversen, H. Madsen, C. Rode, Dynamic modeling of presence of occupants using inhomogeneous Markov chains, *Energy and Buildings* 69 (2014) 213–223.
- [2] S. D'Oca, V. Fabi, S. P. Corngati, R. K. Andersen, Effect of thermostat and window opening occupant behavior models on energy use in homes, in: *Building Simulation*, vol. 7, Springer, 683–694, 2014.
- [3] C. Wang, D. Yan, Y. Jiang, A novel approach for building occupancy simulation, in: *Building simulation*, vol. 4, Springer, 149–167, 2011.
- [4] J. Page, D. Robinson, N. Morel, J.-L. Scartezzini, A generalised stochastic model for the simulation of occupant presence, *Energy and buildings* 40 (2) (2008) 83–98.



Figure A.19: The Apartment Building

- [5] K. Sun, D. Yan, T. Hong, S. Guo, Stochastic modeling of overtime occupancy and its application in building energy simulation and calibration, *Building and Environment* 79 (2014) 1–12.
- [6] F. Haldi, D. Robinson, Interactions with window openings by office occupants, *Building and Environment* 44 (12) (2009) 2378–2395.
- [7] F. Haldi, D. Robinson, On the behaviour and adaptation of office occupants, *Building and environment* 43 (12) (2008) 2163–2177.
- [8] H. B. Gunay, W. O'Brien, I. Beausoleil-Morrison, B. Huchuk, On adaptive occupant-learning window blind and lighting controls, *Building Research & Information* 42 (6) (2014) 739–756.
- [9] J. Tanimoto, A. Hagishima, State transition probability for the Markov Model dealing with on/off cooling schedule in dwellings, *Energy and Buildings* 37 (3) (2005) 181–187.
- [10] U. Wilke, F. Haldi, J.-L. Scartezzini, D. Robinson, A bottom-up stochastic model to predict building occupants' time-dependent activities, *Building and Environment* 60 (2013) 254–264.
- [11] V. Fabi, R. V. Andersen, S. P. Corngati, B. W. Olesen, A methodology for modelling energy-related human behaviour: Application to window opening behaviour in residential buildings, in: *Building Simulation*, vol. 6, Springer, 415–427, 2013.
- [12] Z. Yu, B. C. Fung, F. Haghighat, H. Yoshino, E. Morofsky, A systematic procedure to study the influence of occupant behavior on building energy consumption, *Energy and Buildings* 43 (6) (2011) 1409–1417.
- [13] Z. J. Yu, F. Haghighat, B. C. Fung, Advances and challenges in building engineering and data mining applications for energy-efficient communities, *Sustainable Cities and Society*.
- [14] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, F. Descamps, A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison, *Building and environment* 75 (2014) 67–78.
- [15] J. Zhao, B. Lasternas, K. P. Lam, R. Yun, V. Loftness, Occupant behavior and schedule modeling for building energy simulation through office appliance power consumption data mining, *Energy and Buildings* 82 (2014) 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, I. L. MacDonald, *Hidden Markov models for time series: an introduction using R*, CRC Press, 2009.
- [18] H. Madsen, P. Thyregod, *Introduction to general and generalized linear models*, CRC Press, 2011.
- [19] J. Bloem, H. Madsen, J. Kloppenborg, J. Liisberg, J. Cipriano, G. Mor, 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*, 591–599, 2015.

- [20] M. Pinsky, S. Karlin, An introduction to stochastic modeling, Academic press, 2010.
- [21] H. Hersbach, Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather and Forecasting* 15 (5) (2000) 559–570.
- [22] H. Madsen, Time series analysis, CRC Press, 2007.