Kybernetika

Rajmund Drenyovszki; Lóránt Kovács; Kálmán Tornai; András Oláh; István Pintér

Bottom-up modeling of domestic appliances with Markov chains and semi-Markov processes

Kybernetika, Vol. 53 (2017), No. 6, 1100–1117

Persistent URL: http://dml.cz/dmlcz/147087

KYBERNETIKA — VOLUME *53* (2017), NUMBER 6, PAGES 1100–1117

BOTTOM-UP MODELING OF DOMESTIC APPLIANCES WITH MARKOV CHAINS AND SEMI-MARKOV PROCESSES

Rajmund Drenyovszki, Lóránt Kovács, Kálmán Tornai, András Oláh and István Pintér

In our paper we investigate the applicability of independent and identically distributed random sequences, first order Markov and higher order Markov chains as well as semi-Markov processes for bottom-up electricity load modeling. We use appliance time series from publicly available data sets containing fine grained power measurements. The comparison of models are based on metrics which are supposed to be important in power systems like Load Factor, Loss of Load Probability. Furthermore, we characterize the interdependence structure of the models with autocorrelation function as well. The aim of the investigation is to choose the most appropriate and the most parsimonious models for Smart Grid simulation purposes and applications like Demand Side Management and load scheduling. According to our results most appliance types can be modeled adequately with two states (on/off model) and the semi-Markov process can reproduce the properties of an aggregate load well compared to the original time series. With the price of more parameters of the semi-Markov model compared to identically distributed random sequence and first order Markov chain, it gives better results when the autocorrelation function, Loss of Load Probability and Load Factor are considered.

Keywords: appliance modeling, bottom-up, Markov chain, semi-Markov process, smart grid

Classification: 60J20, 60K15, 60K20

1. INTRODUCTION

Information about electricity load is critical for power systems planning [4], operation and control. Electricity load models are very important, as they contain information on the electricity demand on small time scale (from hours to seconds). Uncertainty is one of the most challenging features of electricity load that can be tackled using probabilistic mathematical models. Markov chains are promising candidates because they can describe states (consumption levels) and transition between these states in a probabilistic manner. In our paper we investigate the applicability of independent and identically distributed random sequences (iid), first order Markov (FOM) and higher order Markov (HOM) chains as well as semi-Markov processes (SMP) for bottom-up (appliance level) load modeling. We use appliance time series from publicly available data sets that con-

DOI: 10.14736/kyb-2017-6-1100

tain fine grained power measurements. We evaluate the goodness-of-fit (GoF) of the different models with respect to the number of states and the order (of Markov chains). However, according to our results, most of the appliances can be modeled by two states (i. e. on-state and off-state), there are certain types of devices where the description with multiple states is more realistic or an additional state can represent for instance stand-by operation state as well. Parsimonious models have advantages, because the number of parameters of the mathematical model is minimal [3]. In this respect FOM is preferred, however it cannot describe long time dependence in time series. First order Markov chains can be extended, in one hand we can increase the order of a Markov chain and in the other hand the expectation of exponential distribution of holding times can be loosened to any arbitrary distribution, resulting in a semi-Markov process. In these cases the Markovian property is lost (i. e. one state depends not only on the previous state) and the number of parameters increase, but we get to models which describe the long time dependence more accurately, hence we get a more realistic appliance load model. The comparison of models with respect to order is based on relative measures like AIC (Akaiken information criterion), BIC (Bayesian information criterion). Based on realistic appliance load models, with the ability to mimic the autocorrelation function (ACF), Loss of Load Probability (LOLP) and Load Factor (LF) well, we can build up appropriate aggregate load profiles. These are important for smart algorithm development (e. g. demand side management [19]) in power systems, especially in Smart Grids [21]. As an example, we refer to our earlier proposed probabilistic demand side management algorithm described in [10].

The paper is organized as follows. In Section 2 we are describing relevant papers with details regarding the applied models from the electricity load modeling literature. Our models (bottom-up aggregate model, iid, Markov chains and semi-Markov processes) and methods (determination of the number of states) are introduced in Section 3, while our results considering number of states, order of Markov chains, ACF, LOLP and LF are described in details in Section 4. Conclusions and plans for future work can be found in Section 5.

2. RELATED WORK

The literature distinguishes two different approaches for load modeling: top-down and bottom-up [6, 7, 20]. The top-down modeling approach works at an aggregated level, typically aimed at fitting a historical time series of national energy consumption. Such models tend to be used to investigate the inter-relationships between the energy sector and the economy at large, and could be broadly categorized as econometric and technological top-down models. The bottom-up approach extrapolates the estimated energy consumption of a representative set of individual houses to regional and national levels.

Often cited bottom-up model is proposed in [13]. The model is based on social random factors (determining appliance stocks and daily consumption level). The authors considered seasonal probability factor for consumption cycles with hourly probability factor for weekdays and weekends. The statistical parameters (pdf and std) of the social random factors were extracted from domestic consumption data. The model can be used to generate realistic domestic electricity consumption data on an hourly basis from a few up to thousands of households.

Differential equations are a natural way to describe the dynamic characteristics of appliances. Three dynamic models with simple structures are investigated in [16] to describe the behavior of a refrigerator. Model parameter identification was based on real measurements. The most complex model (applying two differential equations to describe the warm-up phase and three equations when the compressor is on) models the refrigerator most precisely to be used in intelligent control applications. In [17] a stochastic differential equations (SDEs) model of a domestic freezer is proposed using experimental measurements. The models are estimated by maximum likelihood estimation (MLE). As an application a model predictive control (MPC) is applied for shifting the electricity consumption of a freezer in demand response experiments.

Another promising candidates are methods from the field of artificial intelligence to model residential energy consumption. A machine learning approach is proposed in [18] for online learning of appliance usage from no prior knowledge of operation parameters (duty cycles and times of use). Case studies are presented for typical appliances. Kalman filter realization of Bayesian Filtering is applied for learning the probability of time and extent of use of arbitrary appliances from metering data. In [2] a Neural Network method is proposed to model the residential end-use energy consumption. The paper presents NNs for appliances, lighting, and space-cooling components of the model, with a study of the accuracy of predictions and sample results.

A bottom-up Markov chain Monte Carlo method is described in [12]. The load curves are generated based on behavioral, appliance and climate data. It utilizes a Markov Chain Monte Carlo method to model the occupancy in a household (based on time use survey data). The occupancy pattern with weather data, neighborhood and behavioral characteristics are used to model the appliances. The performance of the model is validated by applying root mean square error (RMSE) with actual measured smart meter data. A hierarchical hidden Markov model (HHMM) framework is introduced in [9] to model home appliances with multiple built-in modes with distinct power consumption profiles (washing machines and dishwashers). An appliance model is presented based on the dynamic Bayesian network representation. A so-called forwardbackward algorithm, performing expectation maximization (EM), is formalized for the HHMM fitting process. Validation on public data shows that the HHMM and proposed algorithm can effectively handle the modeling of appliances with multiple functional modes, as well as better representing a general type of appliances. The authors of the paper [15] proposed a methodology for the modeling of the energy use of small appliances. Seven different modeling approaches were compared. All the models are based on discrete time random (Markov) process and continuous time random process. The validation measures (time series analysis, model accuracy and aggregated energy use) suggest that the modeling of survival in discrete states performs well and is straightforward to integrate with energy simulation programs.

3. APPLIED MODELS AND METHODS

There are many stochastic models that can be used to represent electricity load in Smart Grid applications. We have chosen to use bottom-up appliance electricity load modeling, because our aim was to model appliances in direct control type demand side management algorithms, where individual appliances are influenced (switched on and off, scheduled) to have the desired shape of aggregate load. A possible classification of bottom-up appliance electricity load models are depicted in Figure 1. In one hand the number of states of appliance loads must be determined. Our results show that two states can represent appliance load well (see results in Section 4), hence multistate models are excluded from our further investigations. We consider 3 distinct models: independent and identically distributed random sequences, homogeneous Markov chains and semi-Markov processes. These models are suitable to describe time dependent (or Time-Of-Use) and seasonal behavior if we apply variable probability parameters (time dependent state-transition matrix for Markov chain), though considering short time horizons (usually applied in Demand Side Management and Demand Response applications as being our main modeling targets), we assume stationary time-series exclusively. As an alternative modeling approach, Markov Modulated Poisson Process (MMPP) and its extensions (with arbitrary distributions) can be used to model non-stationary load time series. MMPPs are among the topics of our future research.

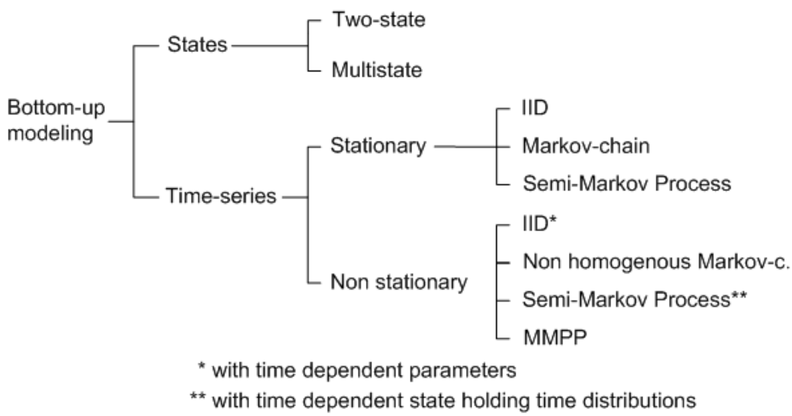


Fig. 1. Classification of bottom-up appliance electricity load modeling.

3.1. Bottom-up aggregate model

A set of appliances represent the load in a given area (e. g. connecting to a bus or transformer), which is a random variable defined as

$$X = \sum_{i=1}^{N} X_i,$$

i. e. X is the sum of X_i individual loads of the appliances and referred to as aggregate load. In this context the individual loads or their statistics are supposed to be known, referred to as bottom-up modeling.

3.2. Determining the number of states

An important question is that how many states are needed in a individual appliance model to best represent the behavior of the appliance, preferably with the least number of parameters. Our approach to determine the number of states is based on the paper [1]. We use k-means clustering algorithm to find k centroids of the representative appliance loads and for the comparison of different state numbers we apply the integral of absolute difference between original and model probability distribution function as a goodness-of-fit measure.

3.2.1. Two state (on/off) modeling

If we reduce the domain of the original discrete random variable $X_j \in \{0, 1, ..., h_j\}$ to on/off values $X_j^{\text{on/off}} \in \{0, h_j\}$ so that we keep the original Expected value $E\{X_j\} = E\left\{X_j^{\text{on/off}}\right\} = m_j$ and maximum value h_j, we get to a random variable $X_j^{\text{on/off}}$ with Bernoulli distribution:

$$\Pr\left(X_j^{\text{on/off}} = 0\right) = 1 - \frac{m_j}{h_j}, \Pr\left(X_j^{\text{on/off}} = 1\right) = \frac{m_j}{h_j} \tag{1}$$

where we need to know only m_j and h_j parameters of the jth user. We will refer this model as on/off max. Alternatively, we can reduce the domain so that we keep the original Expected value and length of operation. This type of model will be referred as On/Off time. In this case we need to recalculate the maximum value:

$$h_j^{\text{new}} = \frac{m_j\left(n_j^{\text{on}} + n_j^{\text{off}}\right)}{n_j^{\text{on}}} \tag{2}$$

where n_j^{on} and n_j^{off} means the number of time instances when X_j is On and Off, respectively. On/off time model is advantageous when our purpose is scheduling, because in this case we need to keep the original length of operation of the appliance j.

3.2.2. Multistate modeling

Multistate models can be represented by general discrete random variables and we can build such models with e.g clustering or requantization. In this paper we focus on two-state models.

3.3. Time-series modeling

In the following sections we introduce the 3 different models that are used in our investigations.

3.3.1. IID model

Let us consider the load time series of a particular appliance as a sequence of independent and identically distributed random variables $X[1], \ldots, X[n]$ taking value (state) from the set of $\{1, \ldots, M\}$ of length n. In this case the aggregate load of all appliances

$(i = 1, \ldots, N)$ at time instant k is the sum of mutually independent random variables $X_i[k]$. Inevitably this is the simplest model which cannot represent time dependencies between time samples, but it is the most tractable from a mathematical point of view.

3.3.2. First order Markov chain

The Markov chain is a stochastic model that represents dependencies between successive observations of a random variable. This model has been used for many decades and it is applicable in many disciplines. In the case of a first order Markov chain the system has finite number of states ($\{1, \ldots, M\}$) and a state depends only on the previous state. At time instant k there is a state transition probability matrix $\mathbf{P}[k]$, where

$$p_{x_1 x_0}[k] = \Pr(X[k] = x_0 | X[k-1] = x_1) \tag{3}$$

where $x_0, x_1 \in \{1, \ldots, M\}$ and $X[k]$ denotes the discrete-time random variable of the time series. In this paper we assume homogeneous Markov chains for which $\mathbf{P}[k] = \mathbf{P}$ stands for all time instances. However, first order Markov chains have more parameters than iid, they are similarly mathematically tractable in many applications. A thorough introduction of Markov chains can be found in [3].

3.3.3. Higher order Markov chain

Higher order Markov chains describe stochastic behavior where the present depends on the last l steps (not only on the previous step). In this case we have an lth-order Markov chain with the transition probabilities

$$p_{x_l \cdots x_2 x_1 x_0}[k] = \Pr(X[k] = x_0 | X[k-1] = x_1, X[k-2] = x_2, \ldots, X[k-l] = x_l) \tag{4}$$

where $x_0, x_1, \ldots, x_l \in \{1, \ldots, M\}$ and $X[k]$ denotes the discrete-time random variable of the time series. Inevitably the advantage of this model is its descriptive power (i. e. it has memory) compared to the First order Markov chain, though the number of parameters increases with the order, which can lead to intractably high number of parameters in the model.

3.3.4. Estimation of the parameters of Markov chains

A Markov chain with any order has $(M-1)$ independent probabilities in each row of the transition matrix. The number of independent parameters to be estimated is equal to $M^l(M-1)$. Given a set of observations, these parameters can be computed as follows. Let $n_{x_l \cdots x_0}$ denote the number of transitions of the type

$$X[k-l] = x_l, \ldots, X[k-1] = x_1, X[k] = x_0 \tag{5}$$

in the data. The maximum likelihood estimate of the corresponding transition probability $p_{x_l \cdots x_0}$ is then

$$\hat{p}_{x_l \cdots x_0} = \frac{n_{x_l \cdots x_0}}{N_{x_l \cdots x_1}} \tag{6}$$

where

$$N_{x_l \cdots x_1} = \sum_{x_0=1}^{M} n_{x_l \cdots x_0} \tag{7}$$

and the log-likelihood of the entire sequence of observations is written

$$LL = \sum_{x_l,\ldots,x_0=1}^{M} n_{x_l \cdots x_0} \log(\hat{p}_{x_l \cdots x_0}). \tag{8}$$

3.3.5. Semi-Markov process

Semi-Markov processes represent a class of stochastic processes that generalize the Markov processes. In the case of a continuous-time Markov process, the state holding times are exponentially distributed, (and respectively for discrete-time Markov chains it is geometrically distributed). In the semi-Markov case, the state holding time distribution can be from any distribution. From this point of view the semi-Markov model is more suitable for many applications, than the Markov one. Semi-Markov processes were first studied by Takacs, Levy and Smith, while Markov renewal processes by Pyke. Consider a state space S and a set of random variables $(X[k], T[k])$. In this case $T[k]$ are the jump times and $X[k]$ are the associated states in the Markov chain. The inter-arrival times are $\tau[k] = T[k] - T[k-1]$ and the sequence $(X[k], T[k])$ is called a Markov renewal process if

$$\Pr(\tau[k] \leq t, X[k] = j | (X[0], T[0]), (X[1], T[1]), \ldots, (X[k-1] = i, T[k-1])) \tag{9}$$

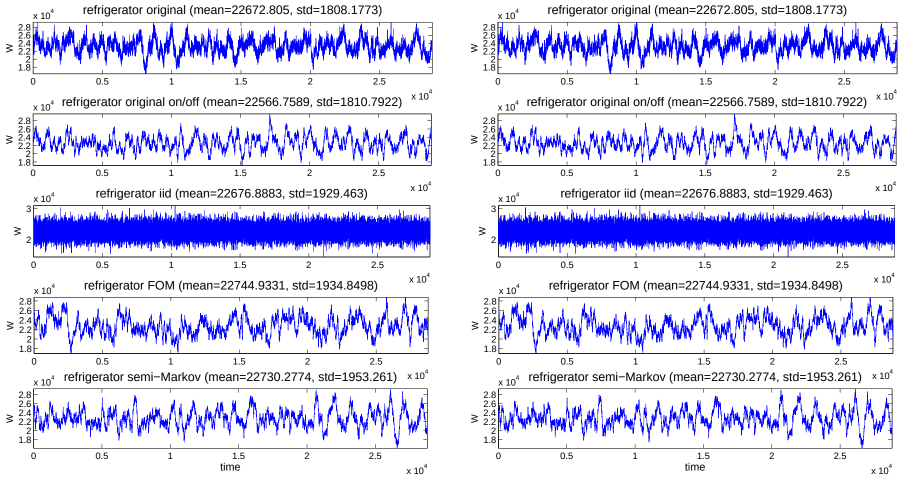
$$= \Pr(\tau[k] \leq t, X[k] = j | X[k-1] = i) \tag{10}$$

$$\forall k \geq 1, t \geq 0, i, j \in \mathrm{S}. \tag{11}$$

We can know define a new stochastic process $Y[t] := X[k]$ for $t \in [T[k], T[k+1])$, where $Y[t]$ is called a semi-Markov process (SMP). The semi-Markov process does not have the Markovian property, so it is not memoryless. Instead the SMP can model arbitrary time holding distribution (from deterministic to stochastic behavior), hence it is a good candidate to use in power system applications to describe load.

4. RESULTS

The comparison of models are based on metrics which are supposed to be important in power systems: Load Factor and Loss of Load Probability. We further characterize the interdependence structure of the models with autocorrelation function as well. Figure 2 and 3 shows time series of the sum of 400 appliances of four types: refrigerator, microwave oven, dishwasher and washer dryer. Original time series are complemented with the on/off versions of the original time series and other three results are generated by the models introduces in the previous sections. The mean of all generated individual appliance time series (iid, FOM and semi-Markov) were matched to the original ones. In the context of bottom-up modeling aggregate consumption time series can be built up by summing individual time series of the appliances. These four Figures show the characteristics of the aggregate (sum of 400 appliances) time series.

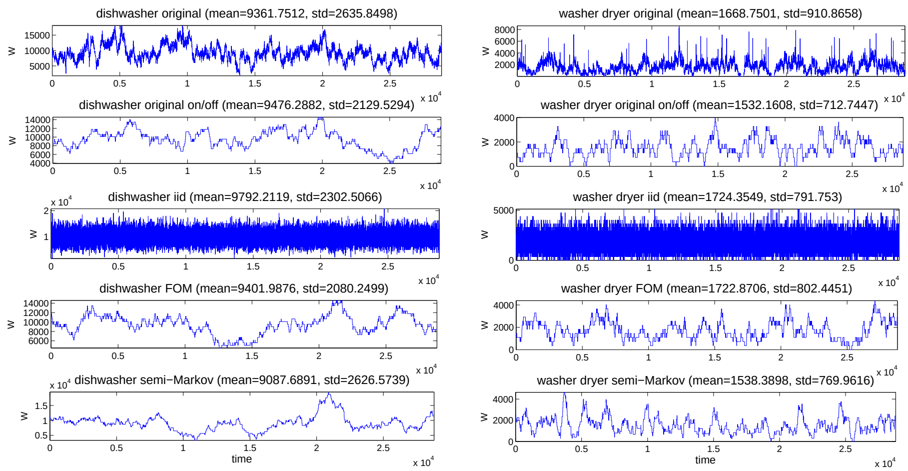


(a) Refrigerator (b) Microwave

Fig. 2. 400 appliances with original on/off, iid, FOM and semi-Markov models.

These four appliance types have different characteristics. Refrigerator operates according to a temperature threshold level, and switches on and off multiple times in every hour. Microwave oven is manually switched on for a short time and it is typically used two or three times a day. Dishwasher operates according to a set of programs and they are used one or two times a day, finally washer dryers are typically operated 3 to 7 times a week and they follow a predefined program as well.

All the models were fitted to measured data coming from the REDD [8] and GREEND [11] datasets. The REDD dataset contains appliance level power data for 6 homes for several weeks with sampling time of 3 seconds, while GREEND dataset provides 3–6 month appliance level power measurements of 8 buildings (9 sensors/home) with 1Hz. In the case of refrigerator, the evaluation of the goodness-of-fit (GoF) shows (Figure 4 (a)) that 3 states would be the most advantageous to use. The integral of absolute difference between original and model pdfs in the case of 2 and 3 states are significantly different (more than 0.1). The reason for three states is that the refrigerator electricity load profile is characterized as three clearly separated load values: off state, a peak instant load when the compressor turns on and the constant load in normal cooling operation. These three phases are far enough to be separated by the k-means clustering algorithm. The load of the microwave oven can be clearly characterized by two distinct states which result is reflected in Figure 4 (b) as well: the difference between GoF value of 2 states and more states is below 0.002.

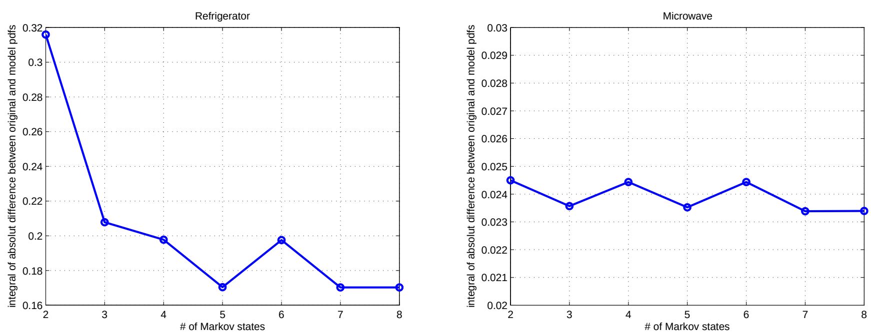


(a) Dishwasher (b) Washer dryer

Fig. 3. 400 appliances with original on/off, iid, FOM and semi-Markov models.

4.1. Number of states

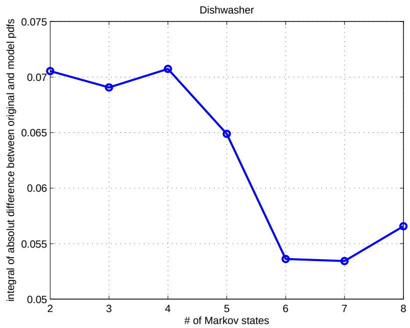
Figure 4 and 5 shows the results of number of states selection.



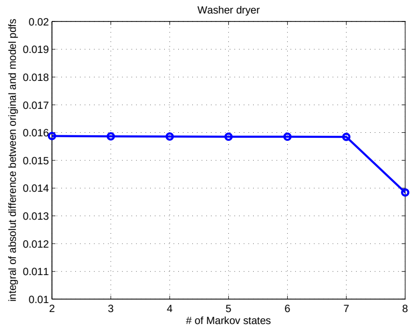
(a) Refrigerator (b) Microwave

Fig. 4. Integral of absolute difference between original and model pdfs.

The results of GoF values of dishwasher and washer dryer appliances are depicted in



(a) Dishwasher



(b) Washer dryer

Fig. 5. Integral of absolute difference between original and model pdfs.

5 (a) and (b) respectively. In the case of dishwasher the difference of GoF values for 2, 3, 4 and 5 states are below 0.005, a small increase (not more than 0.015) can be noticed for 6, 7 and 8 states. The results of washer dryer are similar to ones of microwave and it shows that two states are satisfying to be used for modeling purposes too. Our investigation suggests that two states are suitable in most cases to model appliance load values.

4.2. Order of Markov chains

Two common methods for model comparison are the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC). Both methods use the likelihood ratio statistics with an additional penalty term. The model with the lowest AIC or BIC value is preferred. The Akaiken's Information Criterion is defined by:

$$AIC = -2LL + p \tag{12}$$

The Bayesian Information Criterion (BIC) is defined by:

$$BIC = -2LL + p\log(n), \tag{13}$$

where LL is the log-likelihood of the model (see 8), p is the number of independent parameters and n is the number of components in the log-likelihood. The model achieving the lowest BIC is chosen.

The higher the order the higher the number of parameters are. These two tables contains the number of independent parameters. Measure of goodness such as Log Likelihood (LL) Akaiken Information Criterion (AIC) and Bayesian Information Criterion (BIC) shows no significant differences between the orders. The reason for this result is that the sampling rate is too high compared to the length of state holding times, so a much higher order of Markov chains would be needed to model the behavior of appliances. Alternatively we could decrease the sampling rate, but it would hide the details

	Independence	MC(1)	MC(2)	MC(3)	MC(4)	MC(5)
No. of parameters	1	2	4	8	12	17
LL	-422881	-9988	-9977	-9960	-9949	-9932
AIC	845763	19979	19959	19928	19910	19882
BIC	845775	20004	20009	20028	20060	20095

Tab. 1. Comparison of Higher order Markov models: Refrigerator.

	Independence	MC(1)	MC(2)	MC(3)	MC(4)	MC(5)
No. of parameters	1	2	2	2	3	3
LL	-57083	-2730	-2727	-2723	-2723	-2719
AIC	114167	5463	5456	5449	5449	5442
BIC	114180	5488	5481	5474	5487	5479

Tab. 2. Comparison of Higher order Markov models: Microwave.

in time series data. The conclusion we have reached is that Higher order Markov Chains are not directly applicable for bottom-up load modeling.

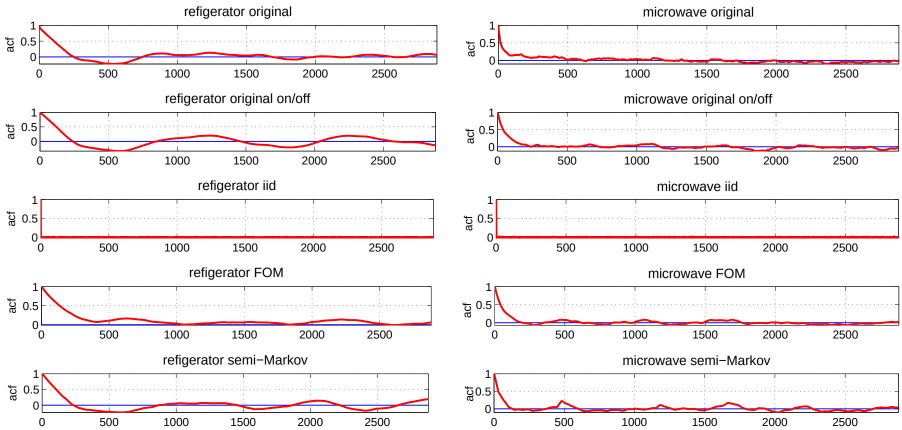
4.3. Autocorrelation function (ACF)

Time series are ordered data containing extra information that can be taken into consideration. Autocorrelation function is one of the tools used to find temporal patterns in time series. More specifically, the autocorrelation function expresses correlation between points with distance of various time lags. The ACF functions of the sum of 400 appliances of refrigerator, microwave oven, dishwasher and washer dryer are depicted in Figure 6 and 7. iid model as containing independent and identically distributed random variables, naturally lacks correlation between consecutive points. This property is shown as a Dirac delta in the ACF (3rd rows in Figures).

The ACF of the sum of 400 appliances are depicted in Figure 6 and 7 comparing iid, FOM and semi-Markov models with the original and on/off version of the original time series. In the case of refrigerator, the autocorrelation structure of the original aggregate time series is well reflected in the ACF of the semi-Markov model, while in the case of microwave, the ACF of FOM is as good as of the semi-Markov. The ACF of semi-Makov model and FOM for dishwasher do not differ too much, while ther is a significant difference in the case of the washer dryer. We can conclude that the autocorrelation structure can well modeled with the semi-Markov model in all cases and in some cases FOM is also suitable. Naturally iid model is not appropriate for his purpose.

4.4. Loss of load probability (LOLP)

In power systems the bus or transformer of a given area has a capacity limit $C_{\max}$ introduced by the physical limitations, or at a given time the supply has a capacity $C < C_{\max}$. Higher demand than powered supply leads to system performance degradation



(a) Refrigerator (b) Microwave

Fig. 6. ACF of 400 appliances with original on/off, iid, FOM and semi-Markov models.

or even to system damage. That is why the event of overconsumption is critical. In mathematical terms we can formulate the event of overconsumption as the probability that a given aggregate load X exceeds the capacity C which is often called as Loss of Load Probability [5] in the reliability assessment literature as well:

$$p = \Pr(X \geq C)$$

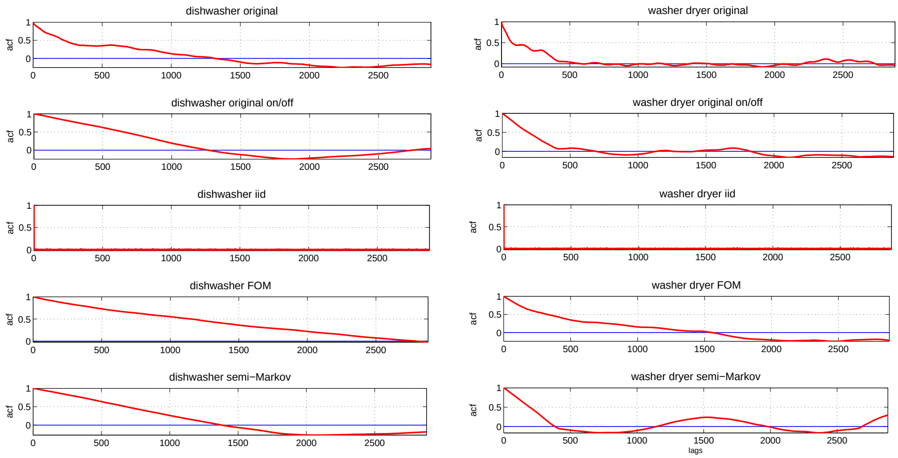
We used an empirical experiment to compare the LOLP values for the different models (Figure 8 and 9). Horizontal axis shows load (in Watts) values from mean to maximum of the original aggregate time series, while vertical axis represents LOLP value in logarithmic scale. The empirical evaluation shows that LOLP values of semi-Markov model are closer to the values of the original compared to the iid and FOM models.

4.5. Load factor (LF)

In power systems the Load Factor (LF) is defined as the average load divided by the maximum load, for a given period of time and it is widely used in the power industry to indicate the efficiency of electricity use [14].

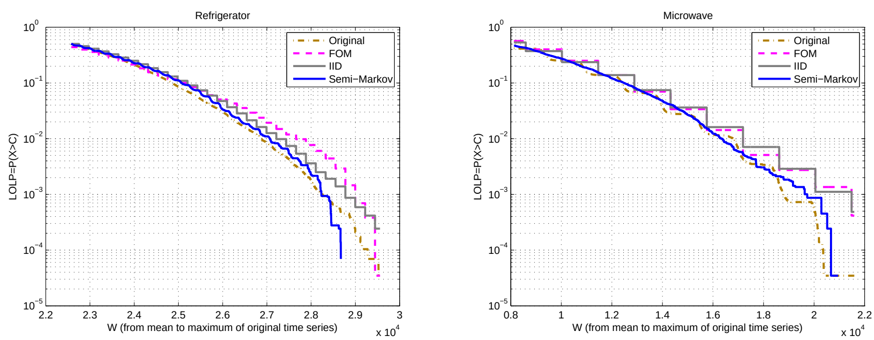
$$LF = \frac{\text{Average load (in given time period)}}{\text{Maximum load (in given time period)}}. \tag{14}$$

However, constant load means that $LF = 1$, its value is always less than one (because average demand is always less than maximum demand). LF can be derived from the



(a) Dishwasher (b) Washer dryer

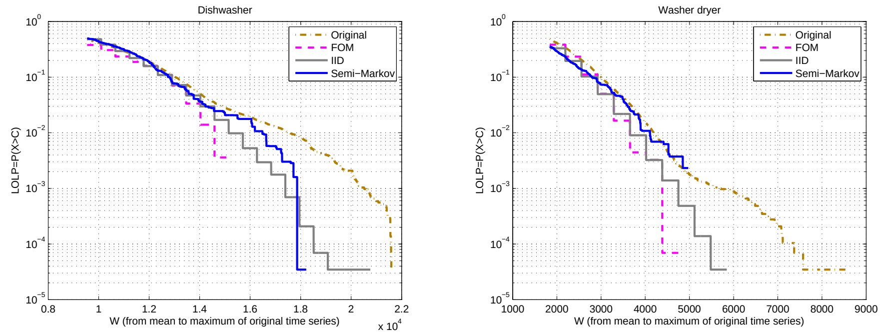
Fig. 7. ACF of 400 appliances with original on/off, iid, FOM and semi-Markov models.



(a) Refrigerator (b) Microwave

Fig. 8. Loss of Load Probability (LOLP) of aggregate of 400 appliances (Original, iid, FOM and semi-Markov models).

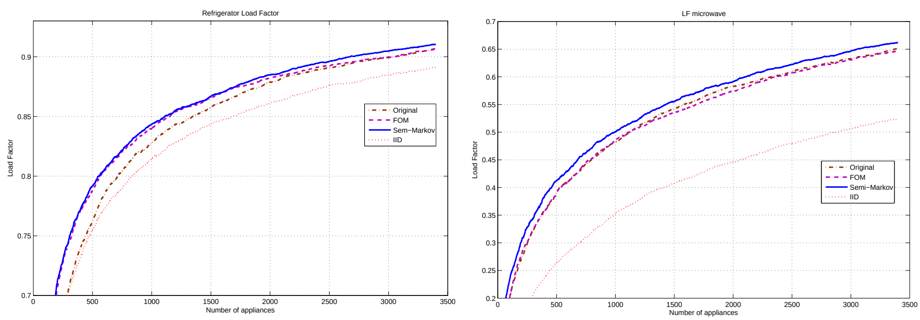
load profile of the specific appliance. A high LF shows almost constant power usage, while low LF means high peaks in demand. For non-peak periods, capacity is idle, which means higher costs for system operators. Additionally electrical rates are designed so that customers with high LF pay less. LF as a metric can also be applicable in load balancing or peak shaving scenarios. As an average behavior, we evaluated the expected



(a) Dishwasher (b) Washer dryer

Fig. 9. Loss of Load Probability (LOLP) of aggregate of 400 appliances (Original, iid, FOM and semi-Markov models).

value of LF as the mean of 30 experiments. Mean LF values are calculated for the sum of up to 3400 appliances (see Figure 10 and 11).



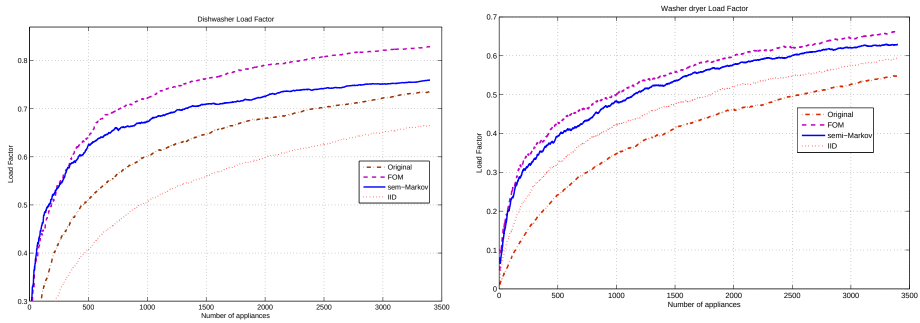
(a) Refrigerator (b) Microwave

Fig. 10. Load Factor for aggregate of 1-3400 appliances with iid, FOM and semi-Markov models.

The advantage of iid on/off (actually it is Bernoulli iid) load model is that we can calculate the expected value of the Load Factor analytically as follows. Let us consider n random variables $Y_1, \ldots, Y_n$ all common Bernoulli distribution. The sum of Bernoulli random variables, $X = \sum Y_i$ have Binomial distribution with cdf:

$$F_X(x) = Pr\left(X < x\right) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i \left(1-p\right)^{n-i} . \tag{15}$$

Now let us have m independent experiments (as an independent sequence) of $X_1, \ldots, X_m$.



(a) Dishwasher (b) Washer dryer

Fig. 11. Load Factor for aggregate of 1-3400 appliances with iid, FOM and semi-Markov models.

In this case we can define the maximum value of the sequence as:

$$\bar{X} = \max\{X_1, \ldots, X_m\} \tag{16}$$

Now we can compute the cdf of $\bar{X}$ as

$$F_{\bar{X}}(x) := Pr\left(\bar{X} < x\right) = Pr\left(X_1 < x, X_2 < x, \ldots, X_m < x\right) = \prod Pr\left(X_i < x\right). \tag{17}$$

If $\bar{X}$ is iid, than

$$F_{\bar{X}}(x) = \left[F_X(x)\right]^m \tag{18}$$

$$E[LF] = \frac{np}{E[F_X(x)]^m}. \tag{19}$$

The Load Factor curve (mean LF values) of the semi-Markov model is an upper bound (more optimistic in LF) of the original LF curve.

5. CONCLUSIONS AND FUTURE WORK

Our paper studied different stochastic models to represent appliance loads. Our results show that in the case of most appliance types, an on/off model is adequate and the semi-Markov process can mimic the properties of an aggregate load well compared to the original time series. With the price of more parameters of the semi-Markov model compared to iid and first order Markov chain, it gives better results when the auto-correlation function, Loss of Load Probability and Load Factor are considered. We are planning to continue our research with multistate and non-stationary models as well. IID model with time dependent parameters, non-homogeneous Markov chains, semi-Markov Process with time dependent holding time distributions and Markov Modulated Poisson Processes are good candidates because they are able to describe the time dependent (or Time-Of-Use) behavior of appliances.

ACKNOWLEDGEMENT

This research is supported by EFOP-3.6.1-16-2016-00006 "The development and enhancement of the research potential at John von Neumann University" project. The Project is supported by the Hungarian Government and co-financed by the European Social Fund. This research was partially supported by Pallas Athena Domus Mentis Foundation. The views expressed are those of the author's (authors') and do not necessarily reflect the official opinion of Pallas Athena Domus Mentis Foundation.

(Received January 24, 2017)

REFERENCES

[1] O. Ardakanian, S. Keshav, and C. Rosenberg: Markovian models for home electricity consumption. In: Proc. 2nd ACM SIGCOMM Workshop on Green Betworking – GreenNets '11, 2011. DOI:10.1145/2018536.2018544

[2] M. Aydinalp, V. I. Ugursal, and A. S. Fung: Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. Applied Energy *71* (2002), 87–110. DOI:10.1016/s0306-2619(01)00049-6

[3] A. Berchtold and A. Raftery: The mixture transition distribution model for high-order Markov chains and non-Gaussian time series. Statist. Sci. *17* (2002), 328–356. DOI:10.1214/ss/1042727943

[4] J. Dickert and P. Schegner: Residential load models for network planning purposes. In: Proc. Modern Electric Power Systems 2010, Wroclaw, pp. 1–6.

[5] R. Drenyovszki, L. Kovacs, I. Pinter, A. Olah, K. Tornai, and J. Levendovszky: Power system reliability assessment for the residential sector based on Large Deviation Theory bounds. In: Proc. EnergyCon 2016, IEEE International Energy Conference, Leuven 2016. DOI:10.1109/energycon.2016.7514106

[6] A. Grandjean, J. Adnot, and G. Binet: A review and an analysis of the residential electric load curve models. Renewable and Sustainable Energy Reviews *16* (2012), 9, 6539–6565. DOI:10.1016/j.rser.2012.08.013

[7] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic: A review of bottom-up building stock models for energy consumption in the residential sector. Building and Environment *45* (2010), 1683–1697. DOI:10.1016/j.buildenv.2010.01.021

[8] J. Z. Kolter and M. J. Johnson: REDD: A public data set for energy disaggregation research. In: Proc. SustKDD Workshop on Data Mining Applications in Sustainability, 2011.

[9] W. Kong, Z. Y. Dong, and D. J. Hill: A hierarchical hidden Markov model framework for home appliance modelling. IEEE Trans. Smart Grid *PP* (2016), 99, 1–1. DOI:10.1109/tsg.2016.2626389

[10] L. Kovacs, R. Drenyovszki, A. Olah, J. Levendovszky, K. Tornai, and I. Pinter: A probabilistic demand side management approach by consumption admission control. Tehnicki Vjesnik – Tehnical Gazette *24* (2017), 1, 199–207. DOI:10.17559/tv-20151021201400

[11] A. Monacchi, D. Egarter, W. Elmenreich, S. D'Alessandro, and A. M. Tonello: GREEND: An energy consumption dataset of households in Italy and Austria. In: Proc. 5th IEEE International Conference on Smart Grid Communications (SmartGridComm 14), Venice 2014. DOI:10.1109/smartgridcomm.2014.7007698

[12] M. Nijhuis, M. Gibescu, and J. F. G. Cobben: Bottom-up Markov Chain Monte Carlo approach for scenario based residential load modelling with publicly available data. Energy and Buildings *112* (2016), 121–129. DOI:10.1016/j.enbuild.2015.12.004

[13] J. Paatero and P. Lund: A model for generating household electricity load profiles. Int. J. Energy Research *30* (2006), 273–290. DOI:10.1002/er.1136

[14] S. N. Palacio, K. F. Valentine, M. Wong, and K. M. Zhang: Reducing power system costs with thermal energy storage. Appl. Energy *129* (2014), 228–237. DOI:10.1016/j.apenergy.2014.04.089

[15] A. Sancho-Tomas, M. Sumner, and D. Robinson: A generalised model of electrical energy demand from small household appliances. Energy and Buildings *135* (2017), 350–366. DOI:10.1016/j.enbuild.2016.10.044

[16] T. Schne, Sz. Jasko, and Gy. Simon: Dynamic models of a home refrigerator. In: Proc. 5th International Conference on Recent Achievements in Mechatronics, Automation, Computer Sciences and Robotics (MACRo 2015), pp. 103–112.

[17] F. Sossan, V. Lakshmanan, G. T. Costanzo, M. Marinelli, P. J. Douglass, and H. Bindner: Grey-box modelling of a household refrigeration unit using time series data in application to demand side management. Sustainable Energy, Grids and Networks *5* (2016), 1–12. DOI:10.1016/j.segan.2015.10.003

[18] B. Stephen, S. Galloway and G. Burt: Self-learning load characteristic models for smart appliances. IEEE Trans. Smart Grid *5* (2014), 5, 2432–2439. DOI:10.1109/tsg.2014.2318375

[19] G. Strbac: Demand side management: Benefits and challenges. Energy Policy *36* (2008), 4419–4426. DOI:10.1016/j.enpol.2008.09.030

[20] L. G. Swan and V. Ismet Ugursal: Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renewable Sustainable Energy Rev. *13* (2009), 1819–1835. DOI:10.1016/j.rser.2008.09.033

[21] Y. Zhang, W. Chen, and W. Gao: A survey on the development status and challenges of smart grids in main driver countries. Renewable Sustainable Energy Rev. *79* (2017), 137–147. DOI:10.1016/j.rser.2017.05.032

Rajmund Drenyovszki, Department of Information Technology, GAMF Faculty of Engineering and Computer Science, John von Neumann University, Izsaki ut 10, H-6000 Kecskemet, Hungary

e-mail: rajmund.drenyovszki@gamf.uni-neumann.hu

Lóránt Kovács, Department of Information Technology, GAMF Faculty of Engineering and Computer Science, John von Neumann University, Izsaki ut 10, H-6000 Kecskemet, Hungary

e-mail: lorant.kovacs@gamf.uni-neumann.hu

Kálmán Tornai, Faculty of Information Technology, Pazmany Peter Catholic University, Prater utca 50/a, H-1083 Budapest, Hungary

e-mail: tornai.kalman@itk.ppke.hu

András Oláh, Faculty of Information Technology, Pazmany Peter Catholic University, Prater utca 50/a, H-1083 Budapest, Hungary
e-mail: olah.andras@itk.ppke.hu

István Pintér, Department of Information Technology, GAMF Faculty of Engineering and Computer Science, John von Neumann University, Izsaki ut 10, H-6000 Kecskemet, Hungary
e-mail: istvan.pinter@gamf.uni-neumann.hu