

Comparison of machine learning methods for estimating energy consumption in buildings

Citation for published version (APA):

Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2014). Comparison of machine learning methods for estimating energy consumption in buildings. In *Proceedings of the 13th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 7-10 July 2014, Durham, United Kingdom* (pp. 1-5). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/PMAPS.2014.6960635>

DOI:

[10.1109/PMAPS.2014.6960635](https://doi.org/10.1109/PMAPS.2014.6960635)

Document status and date:

Published: 01/01/2014

Document Version:

Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

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 the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Comparison of Machine Learning Methods for Estimating Energy Consumption in Buildings

Elena Mocanu*, Phuong H. Nguyen*, Madeleine Gibescu*, Wil L. Kling*,

*Eindhoven University of Technology, Department of Electrical Engineering
5600 MB Eindhoven, The Netherlands

Email: e.mocanu@tue.nl, p.nguyen.hong@tue.nl, m.gibescu@tue.nl, w.l.kling@tue.nl

Abstract—The increasing number of decentralized renewable energy sources together with the grow in overall electricity consumption introduce many new challenges related to dimensioning of grid assets and supply-demand balancing. Approximately 40% of the total energy consumption is used to cover the needs of commercial and office buildings. To improve the design of the energy infrastructure and the efficient deployment of resources, new paradigms have to be thought up. Such new paradigms need automated methods to dynamically predict the energy consumption in buildings. At the same time these methods should be easily expandable to higher levels of aggregation such as neighbourhoods and the power distribution grid. Predicting energy consumption for a building is complex due to many influencing factors, such as weather conditions, performance and settings of heating and cooling systems, and the number of people present. In this paper, we investigate a newly developed stochastic model for time series prediction of energy consumption, namely the Conditional Restricted Boltzmann Machine (CRBM), and evaluate its performance in the context of building automation systems. The assessment is made on a real dataset consisting of 7 weeks of hourly resolution electricity consumption collected from a Dutch office building. The results showed that for the energy prediction problem solved here, CRBMs outperform Artificial Neural Networks (ANNs), and Hidden Markov Models (HMMs).

Keywords—Energy prediction, Stochastic method, Artificial Neural Networks, Hidden Markov Models, Conditional Restricted Boltzmann Machines.

I. INTRODUCTION

Commercial and industrial buildings represent a tremendous amount of the global energy used. A future energy ecosystem is emerging, that connects green buildings with a smart power grid to optimize energy flows between them. This requires prediction of energy consumption in a wide range of time horizons. It is important to predict not only aggregated but to go deep into the individual building level so distributed generation resources can be deployed based on the local forecast. Decomposition of demand forecasting helps analyze energy consumption patterns and identify the prime targets for energy conservation. Moreover, prediction of temporal energy consumption enables building managers to plan out the energy usage over time, shift energy usage to off-peak periods, and make more effective energy purchase plans.

The complexity of building energy behavior and the uncertainty of the influencing factors, such as more fluctuations in demand, make energy prediction a hard problem. These fluctuations are given by weather conditions, the building

construction and thermal properties of the physical materials used, the occupants and their behavior, sub-level systems components lighting or HVAC (Heating, Ventilating, and Air-Conditioning). Many approaches have been proposed aiming at accurate and robust prediction of the energy consumption. In general, they can be divided into two types. The first type of models is based on physical principles to calculate thermal dynamics and energy behavior at the building level. Some of them include models of space systems, natural ventilation, air conditioning system, passive solar, photovoltaic systems, financial issue, occupants behavior, climate environment, and so on. Overall, the numerous approaches depend on the type of building and the number of parameter used. The second type is based on statistical methods. These methods are used to predict building energy consumption by correlating energy consumption with influencing variables such as weather and energy cost. Interested readers are referred to [1] and [2] for a more comprehensive discussion of the application of building energy systems, and more recently reviews [3] and [4]. Moreover, to shape the evolution of future buildings systems there are also some hybrid approaches which combine some of the above models to optimize predictive performance, such as [5]–[8]. Actually, the most widely used machine learning methods for energy prediction are Artificial Neural Networks (ANNs) [9] [10] and Support Vector Machines [11]. Hidden Markov Model (HMM) [12] is other popular stochastic model for time series analyses. This model show good results in different fields, from bio-informatics to stock market and it was not so much investigated in the context of building energy prediction [13].

This paper focuses especially on stochastic methods for energy prediction, by the characterization of load profiles on measured data. Due to the fact that energy consumption can be seen as a time series problem, this paper investigates Conditional Restricted Boltzmann Machine (CRBM) [14], a recent introduced stochastic machine learning method which was used successfully until now to model high non-linear time series (e.g. human motion style, structured output prediction) [15] [16]. Up to our knowledge, this method has never been used in the context of building energy forecasting. The method is compared with the widely used ANNs and HMMs for energy prediction.

The content of this paper is organized as follows. Section II presents the mathematical formalism of the energy prediction problem. Section III describes the formulation and derivation of the mathematical models proposed. In Section IV

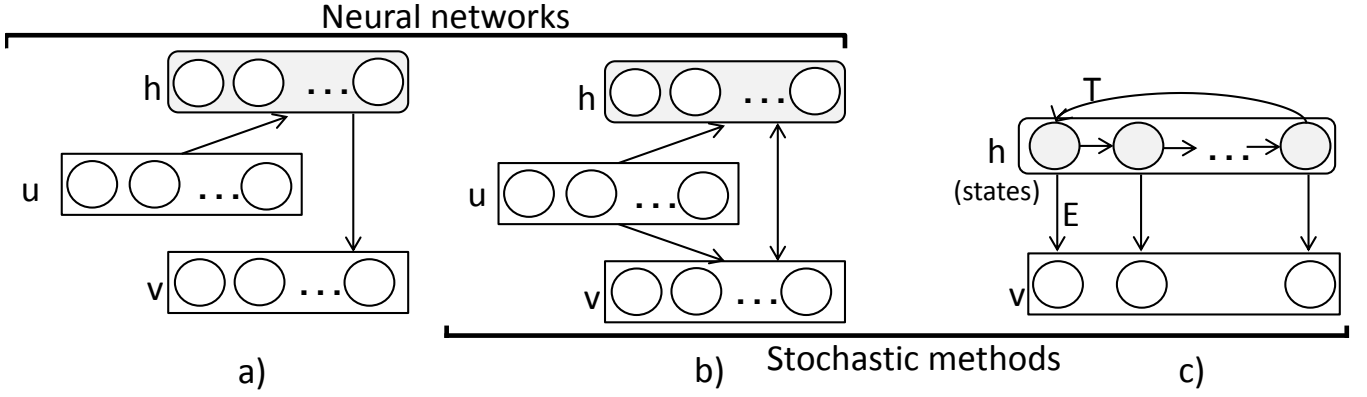


Fig. 1. The general architecture of the models used in this paper to predict the energy consumption: a) Artificial Neural Networks b) Conditional Restricted Boltzmann Machines, and c) Hidden Markov Models. In both types of neural networks u is the conditional history layer (input), h is the hidden layer and v is the visible layer (output).

experimental validation of the methods is shown. Finally, in Section IV, conclusions are drawn and recommendations for future research are given.

II. PROBLEM DEFINITION

The overall technical area of the efficient and effective use of electricity in support of the power systems and customers needs to cover all activities by focusing on advanced end-use efficiency and effective electricity utilization [17], [18]. Modeling and predicting energy consumption in the building can provide valuable information to enable Demand Response or Demand Side Management programs. A simple daily building profile can solve problems that occur frequently, such as load shifting, valley filling or peak clipping. Moreover, these results are used in strategies that encourage the use of electricity more efficiently. Predicting the energy consumption is equivalent to minimize the distance between the real and estimated values.

More formally, let define the following: $i \in \mathbb{N}_l$ represents the index of energy consumption data instances, $t \in \mathbb{N}_T$ denotes time and $\chi \subset \mathbb{R}^d$ represents a d -dimensional feature space.

Given a data set $\mathcal{D}_{Energy} = \{\mathbf{U}^{(i)}, \mathbf{v}^{(i)}\}_{i=1}^l$, where $\mathbf{U}^{(i)} \subseteq \mathbb{R}^{d \times (t-N:t-1)}$, is a d -dimensional input sequence where $t - N : t - 1$ represents a temporal window, $\mathbf{v}_t^{(i)} \subseteq \mathbb{R}^d$ is a multidimensional output vector over the space of real-valued outputs, **determine** $p(\mathbf{V}|\mathbf{\Gamma}; \Theta)$, with $\mathbf{V} \subseteq l \times \mathbb{R}^d$ and $\mathbf{\Gamma} \subseteq l \times \mathbb{R}^{d \times (t-N:t-1)}$ representing the concatenation of all outputs and inputs respectively, and Θ represents the model parameter, such that:

$$\text{Distance} \left(p_{model}(\mathbf{V}|\mathbf{\Gamma}; \Theta) || p_{empirical}(\mathbf{V}|\mathbf{\Gamma}) \right)$$

is minimised. In essence, we aim at solving energy consumption prediction. In the next section, the background knowledge useful to the reader to understand the remaining of the paper is presented.

III. PROPOSED SOLUTION

In this section are described the three methods used in Section IV for energy prediction. Firstly, ANNs and HMMs

are briefly introduced. Then, in the last part of this section the mathematical model of Conditional Restricted Boltzmann Machine is discussed in details.

A. Artificial Neural Networks

Nowadays, Artificial Neural Networks (ANNs) [19] is one of the most widely used solution for energy prediction problem [9] [10]. The general design of ANN is inspired by the model of a human brain. Overall, an ANN is composed by neurons¹, grouped in layers and the connection between them. In our specific case, as depicted in Fig. 1.a the ANN has three layers. More exactly, the \mathbf{u} layer represents the inputs, which encode the last values of the energy consumption, the \mathbf{v} layer which contains the output neurons, and the \mathbf{h} layer which has hidden neurons to learn the characteristics of the time series. The connections between neurons are unidirectional, so that the model is able to compute the output values \mathbf{v} from the inputs \mathbf{u} by feeding information through the network.

B. Hidden Markov models

Baum et all. [12] introduced the mathematical formalism for HMMs to handle sequential data. HMM is investigate in order to extend time-series regression models by the addition of a discrete hidden state variable, which allows changing the parameters of the regression models when the state variable changes its value. The generative HMM [20] used to model our data is depicted in Fig. 1.c. This model consists of a discrete-time and a discrete-state Markov chain, with hidden states (latent variable) $h_t \in \{1, \dots, K\}$. Each latent variable is the representation of a specific combination of $\{hour, day\}$. On this construction is added an observation model $p(\mathbf{v}_t|h_t)$. The joint distribution has the form:

$$\begin{aligned} p(\mathbf{v}_{1:T}|\mathbf{h}_{1:T}) &= p(\mathbf{h}_{1:T})p(\mathbf{v}_{1:T}|\mathbf{h}_{1:T}) \\ &= \left[p(h_1) \prod_{t=1}^T P(h_t|h_{t-1}) \right] \left[\prod_{t=1}^T p(v_t|h_t) \right] \end{aligned}$$

The energy consumption values are continuous and we consider the observation model to be a conditional Gaussian:

$$p(\mathbf{v}_T|h_T = k, \theta) = \mathcal{N}(\mathbf{v}_T|\mu_k, \sigma_k)$$

¹Please note that the neurons can be binary or real values

To predict the future energy values, one has to start from a certain state of the HMM and generate a sequence of following states together with their associated observations drawn from their corresponding probability distribution. To learn the probabilities distribution in the HMM the Baum-Welch algorithm [21] can be used.

C. Conditional Restricted Boltzmann Machines

Conditional Restricted Boltzmann Machines (CRBM) [16] are an extension over Restricted Boltzmann Machines [22] used to model time series data and human activities [15]. Restricted Boltzmann Machines have been applied in different machine learning fields including, multi-class classification [23], collaborative filtering [14], among others. They are energy-based models for unsupervised learning. These models are probabilistic, with stochastic nodes and layers, which make them less vulnerable to local minima [15]. Further, due to their multiple layers and their neural configurations, Restricted Boltzmann Machines possess excellent generalisation capabilities [24]. Formally, a Restricted Boltzmann Machine consists of visible and hidden binary layers. The visible layer represents the data, while the hidden increases the learning capacity by enlarging the class of distributions that can be represented to an arbitrary complexity [15]. In CRBMs models [16] the Restricted Boltzmann Machines model are extended by including a conditional history layer. The general architecture of this model is depicted in Fig. 1.b and the total energy function² is calculate considering all possible interactions between neurons and weights/biases, such as

$$E(\mathbf{v}, \mathbf{h}, \mathbf{u}; \mathbf{W}) = -\mathbf{v}^T \mathbf{W}^{vh} \mathbf{h} - \mathbf{v}^T \mathbf{b}^v - \mathbf{u}^T \mathbf{W}^{uv} \mathbf{v} - \mathbf{u}^T \mathbf{W}^{uh} \mathbf{h} - \mathbf{h}^T \mathbf{b}^h \quad (1)$$

where for each variable a brief explanation is given in Table I. It is worth mentioning that in comparison with ANNs, the

TABLE I. VARIABLES USED IN CRBM

| | |
|---|---|
| $\mathbf{u} = [u_1, \dots, u_{n_u}]$ | represents a real valued vector with all history neurons, where n_u is the index of the last history neuron (input) |
| $\mathbf{v} = [v_1, \dots, v_{n_v}]$ | is a real valued vector collecting all visible units v_i , where n_v is the index of the last visible neuron (output) |
| $\mathbf{h} = [h_1, \dots, h_{n_h}]$ | is a binary vector collecting all the hidden units h_j , with n_h is the index of the last hidden neuron |
| $\mathbf{W}^{vh} \in \mathbb{R}^{n_h \times n_v}$ | represents the matrix of all weights connecting \mathbf{v} and \mathbf{h} |
| $\mathbf{W}^{uv} \in \mathbb{R}^{n_u \times n_v}$ | represents the matrix of all weights connecting \mathbf{u} and \mathbf{v} |
| $\mathbf{W}^{uh} \in \mathbb{R}^{n_u \times n_h}$ | represents the matrix of all weights connecting \mathbf{u} and \mathbf{h} |
| $\mathbf{b}^h \in \mathbb{R}^{n_h}$ | represent the biases for hidden neurons |
| $\mathbf{b}^v \in \mathbb{R}^{n_v}$ | represent the biases for visible neurons |
| τ | represent the iteration |
| α | represent the learning rate |

weights in CRBMs can be bidirectional. More exactly, \mathbf{W}^{vh} is bidirectional. The other weights matrices \mathbf{W}^{uv} and \mathbf{W}^{uh} are unidirectional.

1) *Inference in CRBM*: In CRBMs probabilistic inference means determining two conditional distributions. The first is the probability of the hidden layer conditioned on all the other layers, i.e. $p(\mathbf{h}|\mathbf{v}, \mathbf{u})$, while the second is the probability of the present layer conditioned on the others, such as $p(\mathbf{v}|\mathbf{h}, \mathbf{u})$. Since there are no connections between the neurons in the

same layer, inference can be done in parallel for each unit type, leading to:

$$p(h = 1|\mathbf{u}, \mathbf{v}) = \text{sig}(\mathbf{u}^T \mathbf{W}^{uh} + \mathbf{v}^T \mathbf{W}^{vh} + \mathbf{b}^h)$$

where $\text{sig}(x) = 1/1 + \exp(-x)$, and

$$p(v|\mathbf{h}, \mathbf{u}) = \mathcal{N}(\mathbf{W}^{uvT} \mathbf{u} + \mathbf{W}^{vh} \mathbf{h} + \mathbf{b}^v, \sigma^2)$$

where for convenience σ is chosen to be 1. Probability of the hidden neurons is given by a sigmoidal function evaluated on the total input to each hidden unit and probability of the visible neurons is given by a Gaussian distribution over the total input to each visible unit.

2) *Learning for CRBM using Contrastive Divergence*: Parameters are fitted by maximizing the likelihood function. In order to maximize the likelihood of the model, the gradients of the energy function with respect to the weights have to be calculated. Because of the difficulty of computing the derivative of the log-likelihood gradients, Hinton proposed an approximation method called Contrastive Divergence (CD) [25]. In maximum likelihood, the learning phase actually minimizes the Kullback-Leiber (KL) measure between the input data distribution and the approximate model. In CD, learning follows the gradient of:

$$CD_n = D_{KL}(p_0(\mathbf{x})||p_\infty(\mathbf{x})) - D_{KL}(p_n(\mathbf{x})||p_\infty(\mathbf{x}))$$

where, $p_n(\cdot)$ is the distribution of a Markov chain running for n steps. The update rules for each of the weights matrices and biases can be computed by deriving the energy function with respect to each of these variables (i.e., the visible weights). Formally, this can be written as:

$$\frac{\partial E(\mathbf{v}, \mathbf{h}, \mathbf{u})}{\partial \mathbf{W}^{uh}} = -\mathbf{u}\mathbf{h}^T$$

$$\frac{\partial E(\mathbf{v}, \mathbf{h}, \mathbf{u})}{\partial \mathbf{W}^{uv}} = -\mathbf{u}\mathbf{v}^T$$

$$\frac{\partial E(\mathbf{v}, \mathbf{h}, \mathbf{u})}{\partial \mathbf{W}^{vh}} = -\mathbf{v}\mathbf{h}^T$$

The update equation for the biases of each of the layers, are:

$$\frac{\partial E(\mathbf{v}, \mathbf{h}, \mathbf{u})}{\partial \mathbf{b}^v} = -\mathbf{v}$$

$$\frac{\partial E(\mathbf{v}, \mathbf{h}, \mathbf{u})}{\partial \mathbf{b}^h} = -\mathbf{h}$$

Since the visible units are conditionally independent given the hidden units and vice versa, learning can be performed using one step Gibbs sampling, which is carried in two half-steps: (1) update all the hidden units, and (2) update all the visible units. Thus, in CD_n the weight updates are done as follows:

$$\begin{aligned} \mathbf{W}_{\tau+1}^{uh} &= \mathbf{W}_\tau^{uh} + \alpha (\langle \mathbf{u}\mathbf{h}^T \rangle_{data} - \langle \mathbf{u}\mathbf{h}^T \rangle_{recon}) \\ \mathbf{W}_{\tau+1}^{uv} &= \mathbf{W}_\tau^{uv} + \alpha (\langle \mathbf{u}\mathbf{v}^T \rangle_{data} - \langle \mathbf{u}\mathbf{v}^T \rangle_{recon}) \\ \mathbf{W}_{\tau+1}^{vh} &= \mathbf{W}_\tau^{vh} + \alpha (\langle \mathbf{v}\mathbf{h}^T \rangle_{data} - \langle \mathbf{v}\mathbf{h}^T \rangle_{recon}) \end{aligned}$$

²Please note that the total energy function of CRBM should not be confused with the total energy consumption for building.

and the biases updates are:

$$\begin{aligned}\mathbf{b}_{\tau+1}^v &= \mathbf{b}_{\tau}^v + \alpha (\langle \mathbf{v} \rangle_{data} - \langle \mathbf{v} \rangle_{recon}) \\ \mathbf{b}_{\tau+1}^h &= \mathbf{b}_{\tau}^h + \alpha (\langle \mathbf{h} \rangle_{data} - \langle \mathbf{h} \rangle_{recon})\end{aligned}$$

where τ is the iteration and α is the learning rate.

IV. EXPERIMENTS & RESULTS

To achieve the goal of energy prediction the ANNs, HMMs and CRBM models are evaluated and compared using a set of measured data. In this set the collected data highlights the evolution in time of the *total energy consumption* and *light energy consumption* in the case of an Dutch office building with tree floors and an average number of 35 working persons. This dataset contains 2352 values, collected hourly, over seven

TABLE II. GENERAL CHARACTERISTICS OF DATA SETS

| | #instances | Mean | Standard deviation |
|--------------------------|------------|------|--------------------|
| Lighting consumption | 1176 | 3.71 | 3.97 |
| Total energy consumption | 1176 | 9.54 | 8.24 |

weeks. Some general characteristics based on the gathered data, are detailed in Fig. 2 and Table II. In all experiments

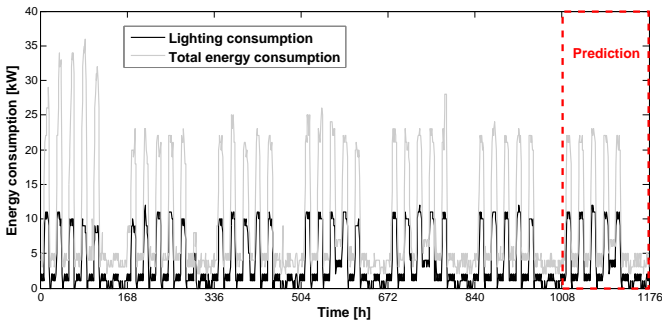


Fig. 2. The time response of the energy consumption in a office building

the aggregated data was separated into the training and test datasets. More exactly the first six weeks were used in the learning phase and the seventh week was used to evaluate the performance of the three methods.

A. Implementation details

We implemented the ANN using the neural network toolbox³ in Matlab and the default settings (i.e. the number of hidden neurons were set to 10 and 1 output neuron). To learn the parameters of the ANN (i.e. the weights between neurons and biases) the network training function was Levenberg-Marquardt optimization algorithm [26]. We used two ANN models in the experiment: the first one was the non-linear autoregressive model with one time series as input (NAR) (i.e. the last week of the energy consumption), and the second one was the non-linear autoregressive model with two time series as input (NARX) (i.e. the last week of the energy consumption plus their corresponding $\{day, hour\}$ states).

The HMM was implemented in Matlab. To predict the energy consumption for one week, we generated a sequence

of 168 states starting from the first hour of the week. With probability one we jumped to the next values. For each state in the sequence we have drawn a sample from the observation distribution which represents the energy consumption for that state.

We implemented the CRBM in Matlab from scratch using the mathematical details described in Section III. The number of hidden neurons was set to 10, the number of output neurons was set to 1, and the learning rate was 10^{-4} . The multi-step prediction of the energy consumption was realized recursively by moving the actual predicted value to the inputs and using it to predict another value in the output layer for an arbitrary number of steps (i.e. 168 steps which represent the numbers of hours in a week).

B. Prediction of energy consumption for lighting

Predicting of the energy consumption to illuminate an office building is directly influenced by human behavior along with many other factors. All of these factors lead to a non-linear time-series. The three models that we proposed to estimate and learn the behavior of this series show very good results. In order to characterize as equidistant as possible the

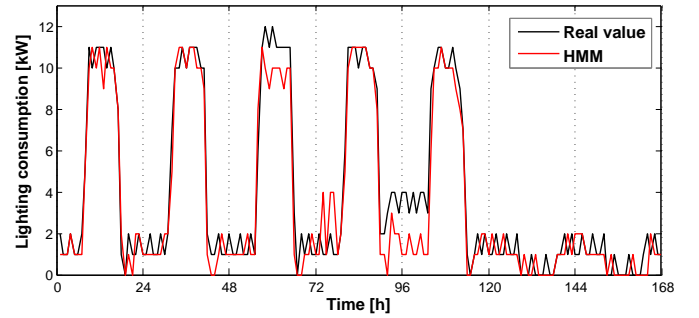


Fig. 3. Prediction of energy consumption for lighting using HMM

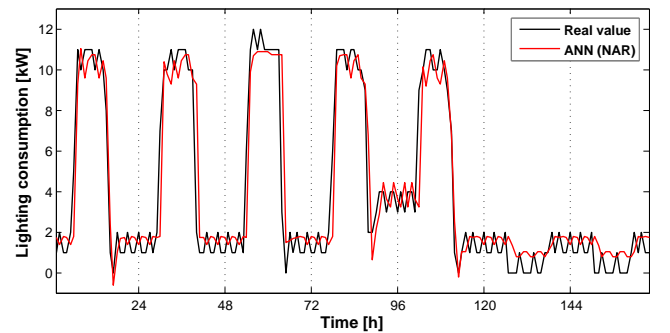


Fig. 4. Prediction of energy consumption for lighting using ANN (NAR)

accuracy of the models used to predict energy consumption we calculated two metrics. Firstly, the prediction accuracy is the ability of a metric to predict with minimum average error and can be evaluated by the root mean square error (RMSE) as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (v_i - \hat{v}_i)^2}$$

³<http://www.mathworks.nl/products/neural-network/>

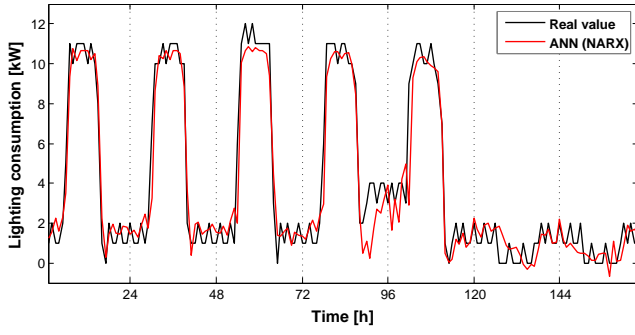


Fig. 5. Prediction of energy consumption for lighting using ANN (NARX)

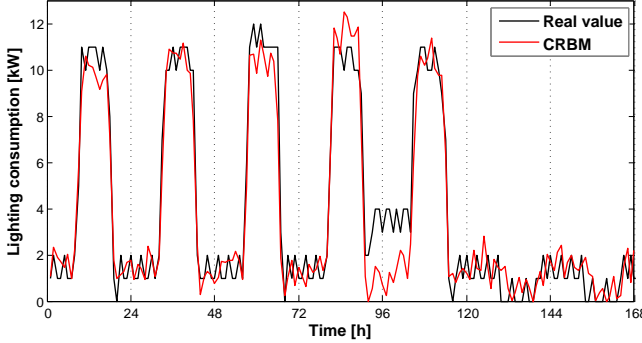


Fig. 6. Prediction of energy consumption for lighting using CRBM

where N represents the total number of data points. Secondly, the correlation coefficient (R) indicating the degree of linear dependence between the real value and the predicted value is define, by:

$$R(u, v) = \frac{E[(u - \mu_u)(v - \mu_v)]}{\sigma_u \sigma_v}$$

where E is the expected value operator with standard deviations σ_u and σ_v . Table III summarizes all the results obtained with ANN-NAR, ANN-NARX, HMM and CRBM to estimate the energy consumption for lighting using these two metrics. Fig. 3, 4, 5, 6 show the detailed prediction results for each method for one week. We can easily observe that CRBM outperforms the other methods.

TABLE III. PREDICTION OF ENERGY CONSUMPTION FOR A WEEK USING HMM, ANN (NAR), ANN (NARX) AND CRBM IN TERMS OF RMSE AND THE CORRELATION COEFFICIENT (R)

| Model | Lighting consumption | | Total energy consumption | |
|------------|----------------------|------|--------------------------|------|
| | RMSE | R | RMSE | R |
| ANN (NAR) | 2.24 | 0.93 | 5.76 | 0.95 |
| ANN (NARX) | 1.22 | 0.96 | 2.52 | 0.98 |
| HMM | 1.23 | 0.95 | 2.55 | 0.95 |
| CRBM | 1.11 | 0.96 | 1.76 | 0.98 |

C. Prediction of total energy consumption in buildings

Predicting the total energy consumption in a building office is even a more difficult problem than the energy consumption for light due to the extra factors which can influence it, such as: weather conditions, HVAC, physical characteristics of the building and so on. In Table III and in Fig. 7, 9, 8, 10 the detailed results are presented for all methods. As in the

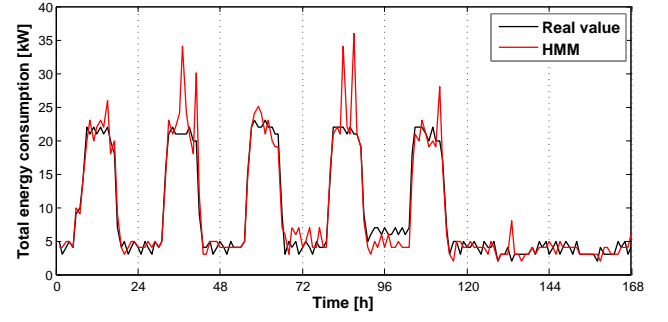


Fig. 7. Prediction of total energy consumption using HMM

previous case, CRBM outperforms all the other methods, but now in the case of total energy consumption prediction the CRBM prediction error is much smaller than the prediction error for the second best model ANN-NARX.

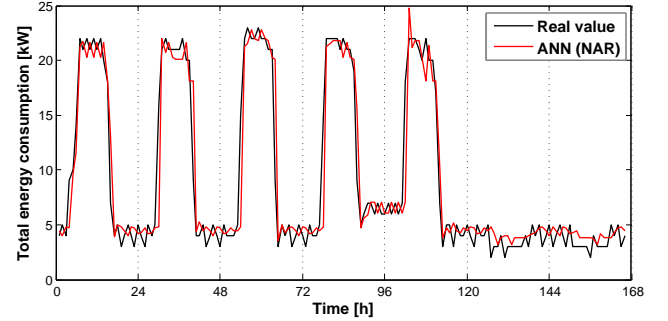


Fig. 8. Prediction of total energy consumption using ANN (NAR)

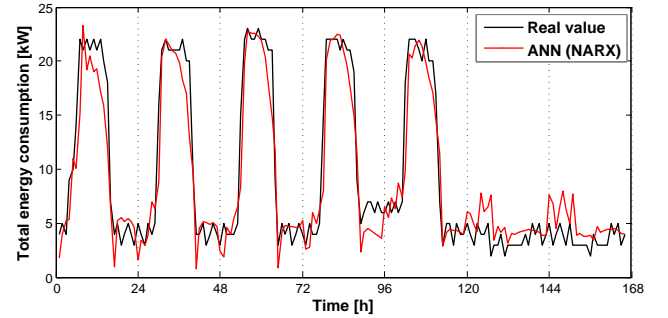


Fig. 9. Prediction of total energy consumption using ANN (NARX)

It is worth mentioning that also in the respect of computational time, CRBM has a slightly small advantage over ANN models. Also HMM is the fastest methods from all with an acceptable error value. However, for this dataset the training time for all methods was of the order of few seconds.

V. CONCLUSION

This paper presented three statistical methods to forecast the energy consumption in an office building over a one week horizon with hourly resolution. Notably, it proposed the use of Conditional Restricted Boltzmann Machines for energy prediction in buildings. The analysis performed showed that

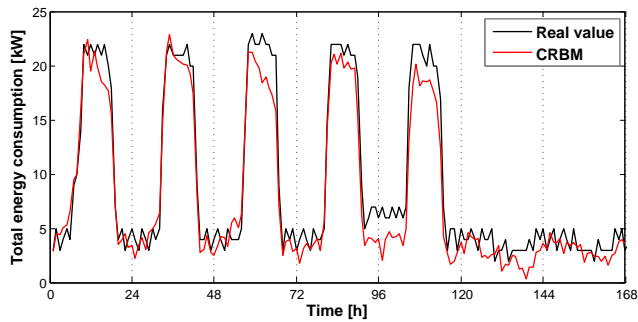


Fig. 10. Prediction of total energy consumption using CRBM

CRBM is a powerful probabilistic method which outperformed the state-of-the-art prediction methods such as Artificial Neural Networks and Hidden Markov Models. Although versatile and successful, these machines come with their own challenges, similar to ANNs. The choice of parameters, such as the number of hidden units and the learning rate must be done carefully. All methods presented showed fast training times, in the order of a few seconds, and are therefore suitable for on-line applications in future building automation systems. As future work, we believe that by adding extra information to the prediction models, such as outside temperature, we can increase the overall accuracy achieved so far.

ACKNOWLEDGMENT

This research has been funded by AgentschapNL - TKI Switch2SmartGrids of Dutch Top Sector Energy. The authors would like to thank to Kropman Installatietechniek B.V. company for providing the dataset.

REFERENCES

- [1] M. Krarti, *Energy Audit of Building Systems: An Engineering Approach, Second Edition*, ser. Mechanical and Aerospace Engineering Series. Taylor & Francis, 2012.
- [2] A. I. Dounis, "Artificial intelligence for energy conservation in buildings," *Advances in Building Energy Research*, vol. 4, no. 1, pp. 267–299, 2010.
- [3] A. Foucquier, S. Robert, F. Suard, L. Stphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," *Renewable and Sustainable Energy Reviews*, vol. 23, no. 0, pp. 272 – 288, 2013.
- [4] H. xiang Zhao and F. Magouls, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586 – 3592, 2012.
- [5] M. Aydinalp-Koksal and V. I. Ugursal, "Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector," *Applied Energy*, vol. 85, no. 4, pp. 271 – 296, 2008.
- [6] L. A. Hurtado-Munoz, P. H. Nguyen, W. Kling, and W. Zeiler, "Building energy management systems : optimization of comfort and energy use," in *Power Engineering Conference (UPEC), 2013 48th International Universities'*, 2013.
- [7] L. Xuemei, D. Lixing, L. Jinhu, X. Gang, and L. Jibin, "A novel hybrid approach of kpca and svm for building cooling load prediction," in *Knowledge Discovery and Data Mining, 2010. WKDD '10. Third International Conference on*, 2010, pp. –.
- [8] B. M. J. Vonk, P. Nguyen, M. Grond, J. Slootweg, and W. Kling, "Improving short-term load forecasting for a local energy storage system," in *Universities Power Engineering Conference (UPEC), 2012 47th International*, Sept 2012, pp. 1–6.
- [9] S. Wong, K. K. Wan, and T. N. Lam, "Artificial neural networks for energy analysis of office buildings with daylighting," *Applied Energy*, vol. 87, no. 2, pp. 551–557, 2010.
- [10] S. A. Kalogirou, "Artificial neural networks in energy applications in buildings," *International Journal of Low-Carbon Technologies*, vol. 1, no. 3, pp. 201–216, 2006.
- [11] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [12] L. E. Baum and T. Petrie, "Statistical inference for probabilistic functions of finite state Markov chains," *Annals of Mathematical Statistics*, vol. 37, pp. 1554–1563, 1966.
- [13] T. Zia, D. Bruckner, and A. Zaidi, "A hidden markov model based procedure for identifying household electric loads," in *IECON 2011 - 37th Annual Conference on IEEE Industrial Electronics Society*, 2011, pp. 3218–3223.
- [14] R. Salakhutdinov, A. Mnih, and G. Hinton, "Restricted boltzmann machines for collaborative filtering," in *In Machine Learning, Proceedings of the Twenty-fourth International Conference (ICML 2004)*. ACM. AAAI Press, 2007, pp. 791–798.
- [15] G. W. Taylor, G. E. Hinton, and S. T. Roweis, "Two distributed-state models for generating high-dimensional time series," *Journal of Machine Learning Research*, vol. 12, pp. 1025–1068, 2011.
- [16] V. Mnih, H. Larochelle, and G. Hinton, "Conditional restricted boltzmann machines for structured output prediction," in *Proceedings of the International Conference on Uncertainty in Artificial Intelligence*, 2011.
- [17] M. Simoes, R. Roche, E. Kyriakides, S. Suryanarayanan, B. Blunier, K. McBee, P. Nguyen, P. Ribeiro, and A. Miraoui, "A comparison of smart grid technologies and progresses in europe and the u.s.," *Industry Applications, IEEE Transactions on*, vol. 48, no. 4, pp. 1154–1162, 2012.
- [18] M. Maruf, L. A. Hurtado-Munoz, P. H. Nguyen, H. L. Ferreira, and W. Kling, "An enhancement of agent-based power supply-demand matching by using ann-based forecaster," in *Power Engineering Conference (UPEC), 2013 48th International Universities'*, 2013.
- [19] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*, 1st ed. Springer, Oct. 2007.
- [20] L. R. Rabiner, "Readings in speech recognition," A. Waibel and K.-F. Lee, Eds. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1990, ch. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition, pp. 267–296.
- [21] L. E. Baum, "An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes," in *Inequalities III: Proceedings of the Third Symposium on Inequalities*, O. Shisha, Ed. University of California, Los Angeles: Academic Press, 1972, pp. 1–8.
- [22] P. Smolensky, "Information processing in dynamical systems: Foundations of harmony theory," in *Parallel Distributed Processing: Volume 1: Foundations*, D. E. Rumelhart, J. L. McClelland *et al.*, Eds. Cambridge: MIT Press, 1987, pp. 194–281.
- [23] H. Larochelle and Y. Bengio, "Classification using discriminative restricted Boltzmann machines," 2008, pp. 536–543.
- [24] Y. Bengio, "Learning deep architectures for AI," *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009, also published as a book. Now Publishers, 2009.
- [25] G. E. Hinton, "Training Products of Experts by Minimizing Contrastive Divergence," *Neural Computation*, vol. 14, no. 8, pp. 1771–1800, Aug. 2002.
- [26] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *SIAM Journal on Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.