Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Energy Procedia 00 (2017) 1–8
Energy Procedia
www.elsevier.com/locate/procedia

# Quantification of energy savings in smart buildings, physics or data?

Aurora González-Vidal, Alfonso Ramallo-González, Antonio F. Skarmeta

**Abstract**

Many efforts from several organisms are focused on the increasement of energy efficiency. It is on the interest of everybody, from particulars to governments, since energy efficiency yields to economical savings for households and companies, reduces greenhouse gas emissions benefiting environment, helps alleviate energy poverty and contributes to growth and jobs. Before implementing an energy efficiency solution, it is neccesary to define the relationship which exists between energy use and the current operating conditions in order to determine energy savings that result form the implementation of the solution.

Traditionally, white box models based on physics equations are used to model systems to predict whole buildings and their sub-systems behaviors, such as their energy consumption and indoor comfort. However, nowadays by means of the Internet of Things we count on vast amounts of data that can be used for knowledge extraction.

We propose a machine learning approach for the creation of an energy consumption prediction model that is used to estimate the energy consumption in normal operation state, i.e., if the system would not have been altered by the energy efficiency experiment. This allows us to compare the predicted for normal operating state and the actual consumption in order to quantify the energy savings.

Our method shows better prediction accuracy than the so-called grey box methods.



*Keywords:*

## 1. Introduction

Energy consumption of buildings in developed countries comprises 20-40% of total energy use and it is above industry and transport in EU and US [1].

In order to adress climate change, together with the use of non-fossil and environmental friendly sources such as solar and wind, cutting energy consumption is also crucial. Reducing energy consumption to the necessary levels so as to keep comfort for buildings users helps to preserve finite resources and lowers the costs for consumers.

When it comes to energy savings, energy management is the process of monitoring, controlling, and conserving energy in a building. Typically this involves the following steps:

- Metering and collecting energy consumption data,
- Proposing ways of saving energy by analysing the data and putting them into practice, and
- Tracking the consumption in order to quantify the gains due to the proposed activity.

The collection of methods and processes used to face the third step, that is to asses the performance of energy efficiency activities by quantifying the gains on efficiency are called Evaluation, Measurement, and Verification (EM&V).

The traditional EM&V methods for determine if a program is generating the expected level of savings are based on linear regression models and they are described in the ASHRAEs Guideline [2].

Regression models need to be typically adjusted in ad hoc manners in order to capture nonlinear behavior, which arises from complex (physical) multivariable interactions between ambient conditions, occupancy levels, and building operating conditions [3].

Regression has always been the standard approach to modeling the relationship between one outcome variable and several input variables, and it can be seen both from a white-box and a black-box point of view. That means, we could use regression for analytical purposes, where a scenario is understood through physics or for data-driven purposes where a scenario is modeled using data alone.

However, with recent advances in sensor and communication technologies, the generation of data about our surrounding environment and ourselves is explosively growing in terms of *velocity*, *variety* and *volume*. This implies that there is more value hidden in the data and as the datasets are generally too large for a p-value to have meaning, predictive data-driven modeling uses other ways to fit a model such as machine learning.

Our proposal for evaluating the gains on energy consumption after an action implementation towards energy efficiency is to take a machine learning approach, where black box models are used in order to predict energy consumption, reducing the cost compared to traditional white box processes which require a level of building engineering expertise that limits scalability.

## 2. Related work

Most of the building energy systems are complex nonlinear systems, which are strongly influenced by weather conditions, building operating modes, and occupant schedules.

Three general categories of building energy forecasting models have been reported in the literature which include white-box (physics-based), black-box (data-driven), and grey-box (combination of physics based and data-driven) modeling approaches [4].

### *2.1. White-box models*

Building structure, systems and equipments need to be considered for this kind of models together with weather conditions. The firsts are usually obtained from design plans, manufacture catalogues or need to be measured in place.

There exist a lot of mature white box simulation engines, that through the combination of mathematical equations simulate the building operation and calculate its energy consumption. Very well-known engines such as EnergyPlus [5] and TRNSYS [6] have been widely used to analyze energy consumption and determine building control and operation scheme [7].

Even though these elaborate simulation tools are effective and accurate, they require detailed information and parameters of buildings, energy systems and outside weather conditions. These parameters, however, are always difficult to obtain, and sometimes they might not be available.

### *2.2. Black-box models*

Black-box models are also known as solely data-driven model. In this case, statistical models are directly applied to capture the relationship between building energy consumption and the inputs: operation and weather data. This type of models need baseline measurements over a certain period of time.

Regression can be used as a data-driven model, however, it is highly more interpretable than machine learning approaches. Machine learning systems figure out how to solve problems with minimal human guidance and once a machine learning algorithm is trained, it can be difficult to understand why it gives a particular response to a set of data inputs. The adaptability of the machine learning models through a self-tuning process, which is different from mathematical models such as regression models, makes accurate decisions without outside expert intervention when unusual perturbations, disturbances, and/or changes in building background conditions occur.

Several data-based modeling techniques have been used for EM&V, including multiple linear regression [8], Gaussian process modeling [3]. In the case of machine learning techniques such as neural networks, support vector machines and their combination [9] and fuzzy logic models [10] are some examples.

### *2.3. Grey-box models*

Grey box models use simplified physical descriptions to simulate the behavior of building energy systems, and then identify important parameters and characteristics using statistical analysis [**?** ].

Data-driven modelling based on grey-box models have been used for many years. As early of 1951 Burnard demonstrated for the first time that resistor-capacitor RC-networks can represent the thermodynamics of buildings accurately [11]. Since then, RC-networks have been used to represent the thermodynamics of buildings. In the early years of building dynamic simulation, this was one of the few ways of representing the thermodynamics of buildings, but even today, programs such as EnergyPlus, include thermal networks on their codes [**?** ] .

In addition to building simulation, grey-box modelling of buildings using RC-networks have been used for the last two decades for Model Predictive Control (MDC). MDC has been used to govern heating and cooling systems of normally large buildings in a way in which the controller can anticipate to the needs of the building via the previous estimation of its thermodynamic features (these normally translated into response times and conductivity of the thermal envelope) [12].

In cases where limited amounts of data are available and the information about the building architecture is partially known, grey models are suitable alternatives for the prediction of electricity consumption [13].

## 3. Methodology

In this section we introduce both a black box and a grey box model based methodology in order to estimate the energy consumption of a building.

### *3.1. Inputs*

Energy forecasting studies that use machine learning are usually intended to predict consumption a priori in order to manage and store the suitable amount of energy, taking into consideration the market prizes and also the necessities of the buildings. However, our approach is different in the sense that our goal is to quantify the energy savings relative to a baseline period due to certain experiment related to efficiency improvement.

This translates into a difference in the inputs that are available for being used. In other scenarios, data concerning energy consumption in previous hours and days is very useful because it is evidently highly correlated with the later consumption [14]. However, we should not use such data since the consumption is altered by the experiment.

On the other side, in other scenarios enviromental data is not available yet (it is the future) and predictions have to be used. When applying prediction for M&V, environmental and occupation variables are usually available and it is not necesary to predict them. The most commonly used weather information is outdoor dry-bulb air temperature.

### *3.2. Modeling*

Our interest lies on weekly quantification of the energy savings. However, daily dynamics are useful since there are patterns that can be found depending on the day of the week. In order to do so, we predict daily energy consumption and then compute the metrics in a aggregated way so that the global quantification is done weekly.

*3.2.1. Black box approach*

Daily aggregate consumption is used as output and we try to capture the relationship between the whole day temperature and the consumption by relating the time series composed by every hours' mean temperature and the daily consumption.

Then, we can use several machine learning models in order to asses which is the best one for our scenarios. The models are generated following the next steps [15]:

- Clean and transform the data: selecting predictive variables such as temperature and day type, deleting outliers
- Aggregate: compute daily consumption, create the time series with the input variables and represent the series in a lower dimension. That is, apply hourly average or other representation and feature selection methods in order to serve as inputs of our models.
- Divide the dataset into train (75 % ) and test (25 %)
- Validation method: 10-fold cross validation and 5 repetitions over the training data set in order to find the hyperparameters of each machine learning algorithm
- Evaluate: apply the algorithm to the test dataset in order to obtain the metric for the model

*3.2.2. Grey box approach*

To make use of the models the set of outputs and inputs have to be defined together with the topology of the system. The most common mathematical representation of lumped parameter models is the state-space representation. The general form for time-invariant models can be written as shown on Eq. 1

$$\begin{cases} x'(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \tag{1}$$

where $x$ is a vector with the states of the model, in our case the temperatures in different nodes of the model, $A$ is a characteristic matrix of the model, $B$ defines the effect of the inputs in the model, and $u$ are the inputs, in our case the temperatures and electric gains. In this formulation, $y$ represents the variables that are measured, in our case electricity. $C$ is the identity matrix; and $D$ is zero in all cases for this work. Using this formulation, every time that a solution had to be evaluated the Octave built in function lsim was used.

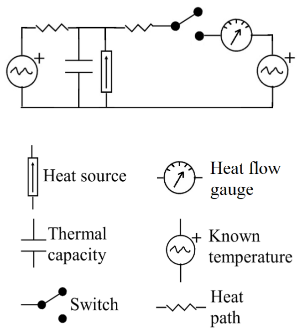


Fig. 1. xxxt

The conditioning system is governed in our case by a thermostat with a timer that turns it on and off. For this reason one need to consider that the RC-network that represents the phenomenon needs to change topologies depending on the operation (on or off) it is for this reason that we have considered a dual-mode RC-network as the one shown in Fig. 1 and previously introduced by Ramallo-Gonzlez on [16].

Once the system was defined, an optimisation algorithm was used to find the values that minimise the RMSE (10 minutes intervals) of the simulated power consumption. To ensure that the data was used adequately, the total electricity consumption was separated onto an un-seasonal component and a seasonal one. The un-seasonal component was used as electric loads and the seasonal component was considered as the heating and cooling load. The building is equipped with a boiler and radiators network (???) that contribute to some of the heating loads. To take into account this effect, the outside temperature on the heating season was risen to a value in which the cooling and heating loads of the building were compensated.

The optimisation method to find the parameters of the model was a simplex. The termination criteria was to get a change on the solution smaller than 0.01 in all parameters. The calculation took approximately 6 minutes on a i5 Intel computer 2.7 running single threaded.

### *3.3. Model Accuracy Metrics*

To assess model accuracy, this work uses two metrics: the mean absolute percentage of error (MAPE) and the coefficient of variation of the room mean squared error (CVRMSE). The MAPE metric has been used in a wide number of electricity prediction studies [17, 18] . It expresses the average absolute error as a percentage and is calculated as follows:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \bar{y}_i|}{y_i} \times 100,$$

where $y_i$ is the real consumption, $\bar{y}_i$ is the predicted consumption and $n$ is the number of observations.

Whereas the CVRMSE has often been used in energy prediction studies [19] . It evaluates how much error varies with respect to the actual consumption mean and is calculated as follows:

$$CVRMSE = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}}{\bar{y}} \times 100,$$

### *3.4. Savings Metrics*

To determine energy savings and uncertainty levels from energy efficiency measures, the IPMVP [13] and ASHRAEs Guideline 14 [2] provide three methods. The one that is suitable for our approach is whole-building metering, since it compares the total energy demand or cost during pre-experiment and post-experiments periods.

How to assess the accuracy and usefulness of whole-building energy models by testing predictions of baseline energy use against actual energy use being the objective to quantify and minimize the uncertainty in reported whole-building savings, which depends on baseline model effectiveness, building predictability, and depth of savings being measured [20].

The predictive baseline models have as an output the metered pre-experiment energy use energy$_{pre}$ and uses the predictors such as environmental conditions inputs$_{pre}$ as inputs of the model. Therefore, the error in reported savings is proportional to the error in the baseline model forecasts.

## 4. Use case

The reference building in which the proposed procedure has been carried out to generate accurate building models is the Chemistry Faculty of the University of Murcia, which is a pilot building for the H2020 project ENTROPY [1].

The data that is used in order to build and train our baseline is 1 year data, from February 2016 to February 2017.

[1]http://entropy-project.eu/

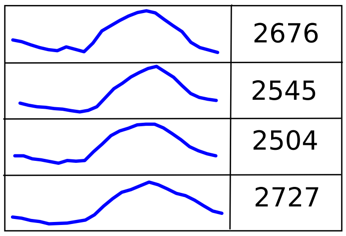


Fig. 2. Daily temperature time series input and consumption output

### 4.1. Black box approach

Our black box methodology is highly versatile with respect to the input data, that is, it allows the addition of variables with minimal effort. In a constructive way, we start by relating the 24 temperature values of each day with the energy consumption of the building (see Fig. 2).

Having introduced the daily temperature time series, we consider that the addition of a categorical variable indicating season is irrelevant. However, the subject building has several features that are typical of educational buildings: the load on weekends is substantially lower than the load on weekdays and there are also differences between the day of the week (mainly Fridays). In those terms, we use analysis of variance (ANOVA) in order to determine wether there exist differences between the consumption of the different days of the week (p-value $= 0.001 < 0.05$) . In a posthoc test, the conclusion is that we can consider that Fridays have a different behaviour than the rest of the days, due to a disminishment on occupation. That way, we can add a dychotomus variable that indicates the kind of the of the week. Weekends and holidays consumption is estimated with the mean of the previous weekends and holidays.

### 4.2. Grey box approach

In the case of our grey box methodology, in order to avoid physically unrealistic results, the data was separated into heating and cooling periods. Any cooling on the heating season or vice versa was made zero.

Once the system was defined, an optimisation algorithm was used to find the values that minimise the RMSE (10 minutes intervals) of the simulated power consumption. To ensure that the data was used adequately, the total electricity consumption was separated onto an un-seasonal component and a seasonal one. The un-seasonal component was used as electric loads and the seasonal component was considered as the heating and cooling load. The building is equipped with a boiler and radiators network that contribute to some of the heating loads. To take into account this effect, the outside temperature on the heating season was risen to a value in which the cooling and heating loads of the building were compensated. The optimisation method to find the parameters of the model was a simplex. The termination criteria was to get a change on the solution smaller than 0.01 in all parameters.

### 4.3. Results

| | | Models | | |
|---|---|---|---|---|
| | | SVM | RF | XGB |
| Daily | CVRMSE | 12.4 | 9 | 11 |
| | MAPE | 7.2 | 6 | 7.3 |
| Weekly | CVRMSE | 6.4 | 5 | 6.2 |
| | MAPE | 5.2 | 4.5 | 5.5 |

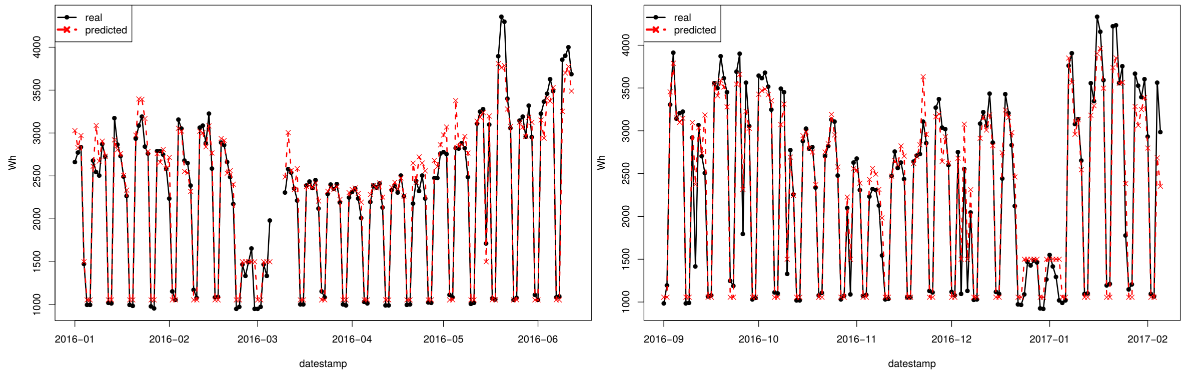


Fig. 3. Daily predictions using RF and real consumption

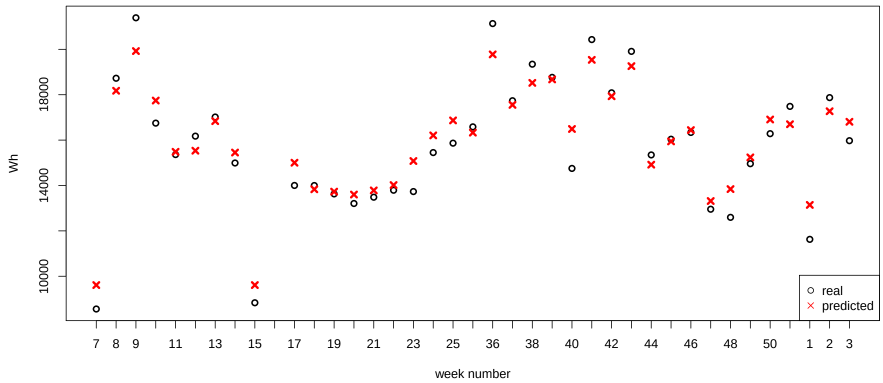


Fig. 4. Weekly predictions using RF and real consumption

## 5. Discussion

## 6. Conclusions and future work

## 7. Acknowledgements

## References


[1] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy and buildings 40 (3) (2008) 394–398.

[2] A. S. Committee, et al., Ashrae guideline 14, measurement of energy and demand savings. atlanta.".

[3] Y. Heo, V. M. Zavala, Gaussian process modeling for measurement and verification of building energy savings, Energy and Buildings 53 (2012) 7–18.

[4] X. Li, J. Wen, Review of building energy modeling for control and operation, Renewable and Sustainable Energy Reviews 37 (2014) 517–537.

[5] D. B. Crawley, L. K. Lawrie, F. C. Winkelmann, W. F. Buhl, Y. J. Huang, C. O. Pedersen, R. K. Strand, R. J. Liesen, D. E. Fisher, M. J. Witte, et al., Energyplus: creating a new-generation building energy simulation program, Energy and buildings 33 (4) (2001) 319–331.

[6] Thermal energy system specialists, `http://www.trnsys.com`, accessed: 2017-09-30.

[7] D. B. Crawley, J. W. Hand, M. Kummert, B. T. Griffith, Contrasting the capabilities of building energy performance simulation programs, Building and environment 43 (4) (2008) 661–673.

[8] M. Braun, H. Altan, S. Beck, Using regression analysis to predict the future energy consumption of a supermarket in the uk, Applied Energy 130 (2014) 305–313.

[9] A. Ahmad, M. Hassan, M. Abdullah, H. Rahman, F. Hussin, H. Abdullah, R. Saidur, A review on applications of ann and svm for building electrical energy consumption forecasting, Renewable and Sustainable Energy Reviews 33 (2014) 102–109.

[10] L. Ciabattoni, M. Grisostomi, G. Ippoliti, S. Longhi, Fuzzy logic home energy consumption modeling for residential photovoltaic plant sizing in the new italian scenario, Energy 74 (2014) 359–367.

[11] G. Burnand, The study of the thermal behaviour of structures by electrical analogy, British Journal of Applied Physics 3 (2) (1952) 50.

[12] D. Coley, J. Penman, Second order system identification in the thermal response of real buildings. paper ii: recursive formulation for on-line building energy management and control, Building and Environment 27 (3) (1992) 269–277.

[13] C. Hamzacebi, H. A. Es, Forecasting the annual electricity consumption of turkey using an optimized grey model, Energy 70 (2014) 165–171.

[14] S. Aman, M. Frincu, C. Charalampos, U. Noor, Y. Simmhan, V. Prasanna, Empirical comparison of prediction methods for electricity consumption forecasting, University of Southern California, Tech. Rep (2014) 14–942.

[15] A. González-Vidal, V. Moreno-Cano, F. Terroso-Sáenz, A. F. Skarmeta, Towards energy efficiency smart buildings models based on intelligent data analytics, Procedia Computer Science 83 (2016) 994–999.

[16] A. P. Ramallo-González, M. Brown, D. A. Coley, Identifying the ideal topology of simple models to represent dwellings.

[17] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, Applied Energy 127 (2014) 1–10.

[18] R. E. Edwards, J. New, L. E. Parker, Predicting future hourly residential electrical consumption: A machine learning case study, Energy and Buildings 49 (2012) 591–603.

[19] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, R. L. Szabados, Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities, IEEE Transactions on Smart Grid 6 (2) (2015) 911–918.

[20] H. Kramer, J. Effinger, E. Crowe, Energy management and information system (emis) software technology assessment: Considerations for evaluating baselining and savings estimation functionality, Pacific Gas and Electric.