A novel data driven approach for the quantification of energy savings in smart buildings

Aurora González-Vidal
Department of Information and Communication Engineering
Computer Science
University of Murcia, Spain
aurora.gonzalez2@um.es

Alfonso P. Ramallo-González
Department of Information and Communication Engineering
Computer Science
University of Murcia, Spain
alfonso.ramallo@um.es

Fernando Terroso-Senz
Department of Information and Communication Engineering
Computer Science
University of Murcia, Spain
fterroso@um.es

Antonio Skarmeta
Department of Information and Communication Engineering
Computer Science
University of Murcia, Spain
skarmeta@um.es

***Abstract*—Many efforts from several organisations are focused on increasing energy efficiency. This is on the interest of everybody, from individuals to governments, since energy efficiency yields to economical savings, to reduce greenhouse gas emissions and to alleviate energy poverty. Buildings are one of the largest consumers of primary energy and its efficiency is an important goal for the achievement of the mentioned goals.**

Currently, the new paradigm of the Internet of Things allow us to count on vast amounts of data that can be used for knowledge extraction of all kinds, one can expect that energy prediction will not be an exception.

This new paradigm that brings large amounts of data, and the advantages on machine learning, have motivated us to test the hypothesis of this paper. This hypothesis supports that prior information that one can find on the physics of the building heat transfer, is now redundant due to the completeness of the data from the system.

We propose a machine learning approach and a grey box model approach to test this hypothesis, the first blind to the physiscs of the problem, and the second heavily influenced by it. The energy consumption prediction models were created with the two approaches. The method has been designed to be used on a project in which the evaluation of energy efficiency interventions is tested. However, the model at this stage was designed to estimate the energy consumption in normal operation state i.e. if the system would not have been altered by any energy efficiency intervention.

Our black box method, which is based on a combination of statistical, machine learning models and on a time series structurisation of the data, shows better prediction accuracy than the so-called grey box methods that include basic physical equations. This proves for this case that the hypothesis of this paper can be rejected i.e. also in this field a data driven approach outperforms more informed methods.

***Keywords*-data-driven models, black box models, grey box models, smart buildings, data analytics**

I. Introduction

Energy consumption of buildings in developed countries comprises 20-40% of their total energy use and it is above industry and transport in EU and US [1], [2].

In order to mitigate climate change, the reduction of energy use together with the use of non-fossil sources such as solar and wind is crucial. Reducing energy consumption on buildings has to be done always maintaining the necessary levels so as to keep comfort for buildings users and to lower their costs to not increase fuel-poverty.

When it comes to energy savings, energy management should be the process of monitoring, controlling, and conserving energy during buildings' normal use. Novel energy feedback systems involve 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 series of methods and processes used to face the third step, that is to asses the performance of energy efficiency interventions by quantifying the gains on efficiency are commonly noted as 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 [3].

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 [4].

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.

In recent years, a new phenomenon is affecting the building sector, and this is the proliferation of smart meters and in home displays, and the trend seems to be on the rise considering that the European Commission has established that 16 Member States will proceed with large-scale roll-out of smart meters by 2020 or earlier [5]. This together with the new developments on Energy Data Infrastructure (such as [6]) has formed the perfect growing soil for the creation of advanced energy feedback strategies for the reduction of energy use in buildings and for the education of building occupants/users [7].

Thanks to that, 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 and comparing it with a grey box approach, that lay between the previous.

II. 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 [8].

A. 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 [9] and TRNSYS [10] have been widely used to analyze energy consumption and determine building control and operation scheme [11].

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.

B. Black-box models

Black-box models are also known as solely data-driven models. 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 [12]. In the case of machine learning techniques such as neural networks, support vector machines and their combination [13], Gaussian process modeling [4], and fuzzy logic models [14] are some examples. In such studies, the way of relating energy consumption and the input variables is one to one, that means, the energy consumption is only related to current indicators such us the instantaneous temperature, humidity and others.

C. 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 [15].

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 [16]. 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 [15] .

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) [17].

This has motivated the investigation of ideal model topologies and methodologies for this search to ensure that the models identified represent accurately the responses of buildings [18]. But also, other works are now focusing on how those methods can be used for their characterisation of a wide variety of buildings in terms of the thermodynamics of the thermal envelope [19].

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 [20].

III. 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.

A. 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 [21]. 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.

B. Proposed models

Our interest lies on weekly quantification of the energy savings. However, daily dynamics are useful since there exist 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.

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 [22]:

- 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

To our knowledge, our method differs from existing methods in the way that the data is introduced. We are relating the input in a time series way with the univariate output. This has been previously done in scenarios where consumption could be used as a predictor [23], something that is not possible in EM&V scenarios.

Also, we avoid the use of change-point models. In a vast amount of the energy efficiency literature, change-point models are used to determine the outdoor air temperatures at which the building transitions from heating to the dead-band and from the dead-band to cooling (like in [24]). However, we cover such effect by using the time series of the whole day temperature.

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.

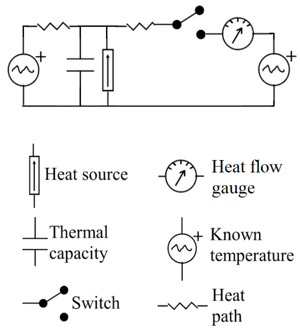

Figure 1: Dual-mode RC network

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-González on [25].

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.

C. Black-box baseline models

Our proposed models are compared with two baseline approaches that are documented in the literature. The first one is a regression-based electricity load model and the second one is a purely machine learning model based that introduces the inputs in a sequential way, relating each

1) Time-of-week-and-temperature (TWT): The algorithm was introduced in [26]. We have chosen it because of its high accuracy, low complexity and low computational cost when compared with the results of other state-of-the-art works used through a wide number of buildings [27].

In this work the predicted load is a sum of two terms: a "time of week effect" that allows each time of the week to have a different predicted load from the others, and a piecewise-continuous effect of temperature.

The main point of this algorithm is the definition of the coefficients that will be estimated using multiple ordinary least square regression. There will be a coefficient for every "time of the week". For example, $24 \times 5 = 120$ coefficients in the case of hourly predictions or $4 \times 24 \times 5 = 480$ in the case of 15 minutes prediction. And also some coefficients related to the temperature. Basically, the range of environmental temperature data is cut into 6 chunks and each of those has an assigned coefficient.

2) Gaussian: In [4] a Gaussian process (GP) modeling framework in order to determine energy savings and uncertainty levels in measurement and verification (M&V) practice is presented. In such work they compare their black box approach to the regular linear regression techniques that are widely exploded on the literature and they state that GP models can capture complex nonlinear and multivariable interactions as well as multiresolution trends of energy behavior thanks to the Bayesian setting under which they are developed.

D. 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 [28], [29] . 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 [30] . It evaluates how much error varies with respect to the actual consumption mean and is calculated as follows:

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

E. 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 [31].

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.

IV. 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.

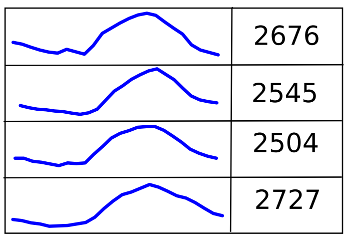

Figure 2: Daily temperature time series input and consumption output

A. 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 day of the week. Weekends and holidays consumption is estimated with the mean of the previous weekends and holidays.

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

The algorithms that we found relevant to use are: Support Vector Regression (SVR), Regression Forest (RF) and Extreme Gradient Boosting (XGB).

SVR works in a similar fashion than Support Vector Machines (SVM). Whereas SVM is a classification technique, SVR fits the optimal curve out of which the training data do not deviate more than a small number ϵ. More specifically, during classification the samples that are close to the margin are penalised even if they are correctly classified, whereas in the regression method an acceptable deviation margin of the samples from the prediction curve is set. The free parameter of this model is C, the penalty parameter of the error term.

Regression Forest is a type of ensemble learning method, where a group of weak models combine to form a powerful model. In Regression Forest multiple regression trees are grown. In order to predict a new observation, each tree gives its own prediction and then the average of them is taken. The algorithm works growing a tree from each random with replacement samples taken of the training. For each node, m_{try} variable are selected at random out of the number of inputs. The best split on these m_{try} is used to split the node. The hyperparameter of regression Forest that we will search for is the number of random variables to take into account on every split: m_{try}.

XGB is built on the principles of gradient boosting and designed for speed and performance (extreme). Gradient Boosted Regression is a technique that generates a prediction through an ensemble of weak prediction models, decision trees in our case. The concept is to sequentially build the model by fitting a weak prediction model on the weighted training data set, where the higher weights are assigned on samples that were previously difficult to predict.

The free parameters of this model are the maximum depth limit of number of nodes in the tree, the minimum number of samples required to split an internal node and the learning rate by which the contribution of each tree is shrunk.

For the sake of comparison, we are using also the Gaussian process for modeling the instantaneous consumption prediction using the current input values (mean daily temperature for daily cosumption or mean hourly temperature for hourly consumption).

B. 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.

C. Results

The prediction metrics are summarised on Table I. The first three methods: SVR, RF, XGB are the ones used for the black box model approach. As it can be seen, they return the best results compared to the Gaussian method, that is applied in a more traditional way, that is relating the instantaneous consumption measurement with the instantaneous inputs measurements and also with our grey box model approach.

Between the three black box mehtods, Random Forest is the one that stands out getting a CVRMSE of 9 and 5 % and a MAPE of 6, 4.5 % for the daily and weekly predictions respectively. We have plotted the prediction vs the real daily consumption in Fig. 3 and weekly in Fig. 4 .

Table I: Metrics

		Models					
		SVR	RF	XGB	TWT	Gauss	Grey
Daily	CVRMSE	12.4	9	11	14.9	17.45	33.57
	MAPE	7.2	6	7.3	12.3	15.01	43.02
Weekly	CVRMSE	6.4	5	6.2	11.1	16.3	19.53
	MAPE	5.2	4.5	5.5	9.4	12.3	15.48

V. Discussion

As previously mentioned, this paper aims to evaluate if the fact that the RC-networks on the grey-box inverse modelling contains basic information about the physical system (i.e. how the thermodynamics of the building should be) deliver any advantage for energy consumption prediction of a smart building with measurement and verification purposes over our proposed black-box method in which no prior high level information about the system is included (i.e. they are blind to the physics of the problem) and that combines statistical analysis and machine learning techniques.

We have shown that using the daily temperature time series we are able to capture the behaviour of the people better than if instantaneous values are used for predicting consumption (Gaussian) or if we use a barrage of coefficients in linear regression for modelling each little part of the day (TWT). The drawbacks of the baseline data-driven models are:

- They neglect the correlation between timestamps. TWT creates artificial features in order to model each moment independently. However, we understand it as time series and are considering in a indirect way all the interactions that building and temperature experience towards consumption.
  Regression requires an assumption of normality. Also, additional assumptions for regression are that the mean of the error term is equal to zero, and that the error term has equal variance for different levels of the input or independent variables. While the assumption of zero mean is almost always satisfied, the assumption of equal variance is complicated to get.
-

That way, we have created a methodology where the combination of time series and current state variables is possible and provides a wide variety of techniques related to time series that can be used to improve the results of the analysis.

VI. Conclusions and future work

A very important characteristic of the black box approach is its generality. Even if we would have obtained similar results with the grey box models, black box models are clearly more generalizable than the previous. That way, a future via of research consists on the application of the same models to several buildings that share some characteristics. That is, they belong to the same environment: different faculties of a university (where students and professors behave similarly between buildings), several houses of a neighbourhood or different malls of the same city.

Reducing the total length of time required for EM&V is key to scaling the deployment of efficiency projects in general, and reducing overall costs [32]. That is why a transfer learning approach has to be considered in future studies in order to reduce the quantity of data that needs to be collected for creating a reasonable building model.

Also, a wider study on the way of introducing the time series to the algorithms should be done, that means, applying time series segmentation and representation techniques for finding more representative ways of introducing the data.

Acknowledgements

This work has been partially funded by MINECO TIN2014-52099-R project (grant BES-2015-071956) and ERDF funds, by the European Commission through the H2020-ENTROPY-649849 EU Project. Ramallo-Gonzlez would like to thank CARM via Fundacin Sneca for the program Saavedra Fajardo, project number: 20035/SF/16.

References

[1] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and buildings*, vol. 40, no. 3, pp. 394–398, 2008.

[2] "Monthly energy review," *U.S. Energy Information Administration*, September, 2017.

[3] A. S. Committee *et al.*, "Ashrae guideline 14, measurement of energy and demand savings. atlanta."," 2002.

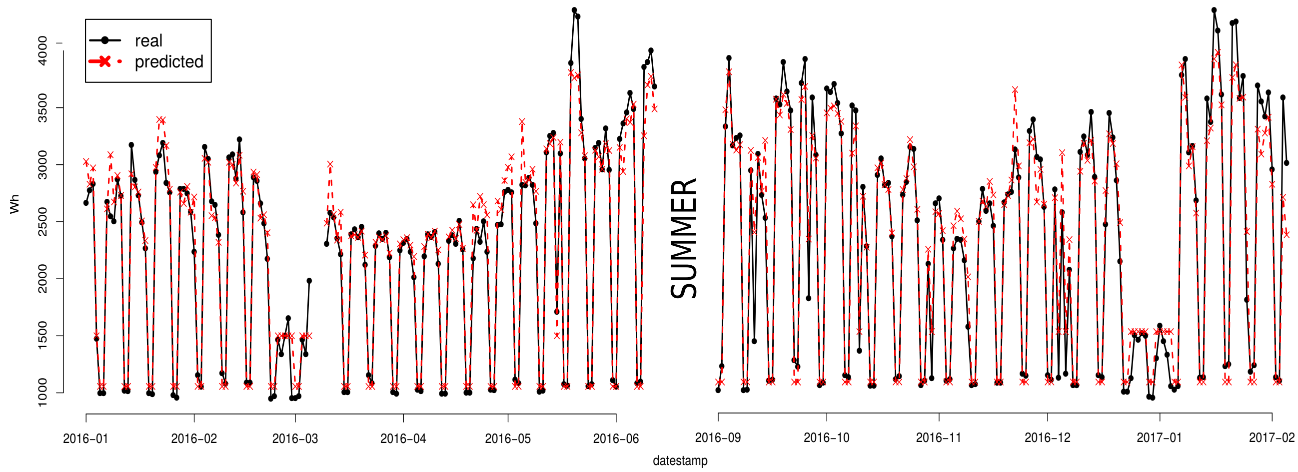

Figure 3: Daily predictions using RF and real consumption

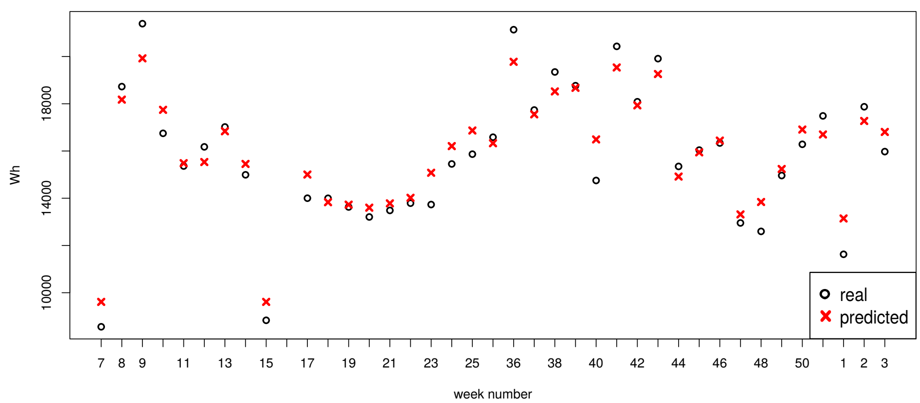

Figure 4: Weekly predictions using RF and real consumption

[4] Y. Heo and V. M. Zavala, "Gaussian process modeling for measurement and verification of building energy savings," *Energy and Buildings*, vol. 53, pp. 7–18, 2012.

[5] EC, "Benchmarking smart metering deployment in the eu-27 with a focus on electricity," *Report from the Commission*, 2014.

[6] F. Terroso-Saenz, A. González-Vidal, A. P. Ramallo-González, and A. F. Skarmeta, "An open iot platform for the management and analysis of energy data," *Future Generation Computer Systems*, 2017.

[7] N. Mogles, I. Walker, A. P. Ramallo-González, S. Lee, J.and Natarajan, J. Padget, E. Gabe-Thomas, T. Lovett, S. Ren, G.and Hyniewska, E. O'Neill, R. Hourizi, and D. Coley, "How smart do smart meters need to be?" *Building and Environment*, no. 125, pp. 439–450, 2017.

[8] X. Li and J. Wen, "Review of building energy modeling for control and operation," *Renewable and Sustainable Energy Reviews*, vol. 37, pp. 517–537, 2014.

[9] 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*, vol. 33, no. 4, pp. 319–331, 2001.

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

[11] D. B. Crawley, J. W. Hand, M. Kummert, and B. T. Griffith, "Contrasting the capabilities of building energy performance simulation programs," *Building and environment*, vol. 43, no. 4, pp. 661–673, 2008.

[12] M. Braun, H. Altan, and S. Beck, "Using regression analysis to predict the future energy consumption of a supermarket in the uk," *Applied Energy*, vol. 130, pp. 305–313, 2014.

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

[14] L. Ciabattoni, M. Grisostomi, G. Ippoliti, and S. Longhi, "Fuzzy logic home energy consumption modeling for residential photovoltaic plant sizing in the new italian scenario," *Energy*, vol. 74, pp. 359–367, 2014.

[15] A. F. Handbook, "American society of heating, refrigerating and air-conditioning engineers," *Inc.: Atlanta, GA, USA*, 2017.

[16] G. Burnand, "The study of the thermal behaviour of structures by electrical analogy," *British Journal of Applied Physics*, vol. 3, no. 2, p. 50, 1952.

[17] D. Coley and 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*, vol. 27, no. 3, pp. 269–277, 1992.

[18] P. Bacher and H. Madsen, "Identifying suitable models for the heat dynamics of buildings," *Energy and Buildings*, vol. 43, no. 7, pp. 1511–1522, 2011.

[19] A. P. Ramallo-González, M. Brown, E. Gabe-Thomas, T. Lovett, and D. A. Coley, "The reliability of inverse modelling for the wide scale characterization of the thermal properties of buildings," *Journal of Building Performance Simulation*, pp. 1–19, 2017.

[20] C. Hamzacebi and H. A. Es, "Forecasting the annual electricity consumption of turkey using an optimized grey model," *Energy*, vol. 70, pp. 165–171, 2014.

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

[22] A. González-Vidal, V. Moreno-Cano, F. Terroso-Sáenz, and A. F. Skarmeta, "Towards energy efficiency smart buildings models based on intelligent data analytics," *Procedia Computer Science*, vol. 83, pp. 994–999, 2016.

[23] Z. Ruijin, P. Yiqun, H. Zhizhong, and W. Qiujian, "Building energy use prediction using time series analysis," in *Service-Oriented Computing and Applications (SOCA), 2013 IEEE 6th International Conference on*. IEEE, 2013, pp. 309–313.

[24] J. K. Kissock, T. A. Reddy, and D. E. Claridge, "Ambient-temperature regression analysis for estimating retrofit savings in commercial buildings," *Journal of Solar Energy Engineering*, vol. 120, no. 3, pp. 168–176, 1998.

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

[26] J. L. Mathieu, P. N. Price, S. Kiliccote, and M. A. Piette, "Quantifying changes in building electricity use, with application to demand response," *IEEE Transactions on Smart Grid*, vol. 2, no. 3, pp. 507–518, 2011.

[27] J. Granderson, S. Touzani, C. Custodio, M. D. Sohn, D. Jump, and S. Fernandes, "Accuracy of automated measurement and verification (m&v) techniques for energy savings in commercial buildings," *Applied Energy*, vol. 173, pp. 296–308, 2016.

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

[29] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy and Buildings*, vol. 49, pp. 591–603, 2012.

[30] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, 2015.

[31] H. Kramer, J. Effinger, and E. Crowe, "Energy management and information system (emis) software technology assessment: Considerations for evaluating baselining and savings estimation functionality," *Pacific Gas and Electric*, 2013.

[32] J. Granderson, P. N. Price, D. Jump, N. Addy, and M. D. Sohn, "Automated measurement and verification: Performance of public domain whole-building electric baseline models," *Applied Energy*, vol. 144, pp. 106–113, 2015.