Model free approach based on ioT data analytics for energy efficiency in smart environments

Aurora González, Victoria Moreno and Antonio F. Skarmeta
Department of Information and Communications Engineering
Computer Science Faculty, University of Murcia, Spain
{aurora.gonzalez2, mvmoreno, skarmeta}@um.es

Abstract—The paradigm of Internet of Things brings a lot of opportunities to benefit all data sensed by the multitude of sensors deployed around us to extract useful knowledge from them in order to understand the specific behaviour of relevant parameters. Following this approach, in this work we describe our model-free procedure from the field of intelligent data analytics that uses the captured sensor data in the context of smart buildings as inputs of mathematical models, with the goal of getting energy control strategy improvements. We study various data-driven techniques for contextual identification of patterns from literature, and from them we select the best performing methods as inputs into our modelling strategy for saving energy.

Keywords—Energy Efficiency; Smart Buildings; Predictive Models

1 Introduction

One of the highest energy consumers in the world are buildings, both residential and commercial, representing between 20% and 40% of the total energy in developed countries [1]. During the last years, reducing the carbon footprint on a global scale and ensuring energy efficiency of buildings are goals with high priority in the fields of building engineering and energy policy.

Focusing on what is happening in Europe, European Commission proposed some years ago the Directive about *"Energy Performance of Buildings"* (2010/31/EU) [2]. This Directive proposes measures to ensure the energy efficient usage of appliances of buildings like lighting, boilers and, especially, heating, ventilation and air conditioning (HVAC) systems.

In this context, the integration and development of systems based on Information and Communication Technologies (ICT) and, more specifically, the Internet of Things (IoT) [3], are important enablers of a broad range of applications, both for industries and the general population, helping to make smart buildings a reality.

In this work we analyze the main factors impacting the energy consumption, taking into account the operational and commercial constraints of the building of interest. Our approach to reducing energy usage in building operations is based on the understanding of how energy is consumed in the building and anticipate its response to make control decisions. For this, we derive predictive models capturing the influence of building characteristics and usage patterns. We compare various data-driven predictive methods in terms of their accuracy when they are applied to modeling our building data. The best performing methods will serve to generate the final models which will be the inputs into a suitable optimization strategy for deriving automatic decisions.

The structure of this paper is as follows: Section 2 presents the two main approaches which are proposed in the literature to implement energy building management systems based on simulated and real data. Section 3 describes our methodology to generate building models which can be used to design the strategy to optimize energy consumption in buildings. Section 4 presents two examples of model generated following the methodology proposed.

Finally, Section 5 gives some conclusions and an outlook of future work.

2 Approaches in Energy Saving Modeling

There exist two broad categories of energy savings approaches for the operational phase of a building that benefit from IoT:

- Simulation based methods requiring physical models of the buildings under study (e.g. [4]–[6]).
- Data-driven approaches from the field of Artificial Intelligence (AI) that adapt regression models to captured sensor data (e.g. [7]–[9]).

The former category requires access to building planning data, e.g. regarding the materials in use for insulation. This can become prohibitively expensive to create for legacy buildings. Typically, simulation methods require access to sensor data for calibration purposes and often have high computational cost [10]. These approaches have difficulties in modeling non-uniform user behavior [11].

Data-driven approaches can be applied directly in settings where sensor data is available and reflect real-life building usage in a realistic way. Typically they accommodate the user data during the learning phase implicitly [12]. In literature both approaches are applied to derive control strategy improvements. While simulation based approaches can also be used to study the potential effects of refurbishment measures such as the effect of replacing windows, data-driven approaches are limited in their predictive capabilities in that regard.

In this paper, and as part of the EU Entropy project [13], we focus on the data-driven control strategy. This approach is applied to get energy usage improvements in the Faculty buildings of the University of Murcia (UMU), Spain, used as pilots for experimentation in the frame of the EU Entropy project. The reason to apply a data-driven approach instead of using simulation tools is because these buildings are totally sensorized and a huge amount of data are already available to be analyzed. These data can be used to extract interesting usability patterns of the buildings, for example the extraction of the energy consumption patterns associated to thermal comfort. Furthermore, an automation platform is installed to integrate all this sensed information as well as to carry out the control of the different electrical devices connected to the platform. This is the Open Data platform, which is derived from some improvements and expansions made over the initial platform presented as Domosec [14], which was developed by the Dept. of Information and Communication Engineering of UMU.

3 Model-free Methodology

In this section we describe our model-free methodology from the field of intelligent data analytics that uses the captured sensor data in the context of smart buildings as inputs of mathematical models, with the goal of getting energy control strategy improvements. The main idea is to study various data-driven techniques for contextual identification of patterns from literature, and from them we select the best performing methods - according to the specific characteristics of the building under analysis - as inputs of the optimization strategy designed for saving energy.

3.1 Pre-proccesing of data

The dataset collected by the device sensors deployed to sense the variables identified as affecting energy consumption in buildings should be cleaned according to the detection of outliers and missing values. Once data are cleaned, correlations between inputs and outputs of the models must be calculated to ensure that, the parameters selected because it is assumed that they are involved in the energy consumption of the target building, are actually affecting.

3.2 Data modeling

When studying the different aspects affecting energy consumption and indoor temperature trends in buildings, we follow a general process to generate the predictive models which will be used later as base of the optimization strategy for the operation of the heating system.

1) Feature extraction. Based on the dataset obtained after preproccesing, compact representations of the inputs, named *features*, are extracted. These features are used later for carrying our the estimation.
2) Train and test data. Partitioning of input feature vectors into training data set (75%) and test data set (25%).
3) Normalization. All values in the dataset are normalized. The resulting values are in the [0,1] range for every feature extracted form the initial dataset.
4) A common technique applied to data is the transformation of the data space using the so called Principal Components Analysis (PCA) [15]. PCA is a widely used technique for reducing dimensionality, identifying the directions in which the variance of the observations is accumulated.
5) Training phase. The regression technique is trained using the 10-fold cross validation and 5 repetitions over the training data set. During this phase different values are used for the setting of the regression techniques (hyperparameters), obtaining different results for each one of them.
6) Evaluation metric: RMSE (Root-Mean-Square Error) and R-Squared. The formula yields the values in the same units as the output of the estimators so the results can be interpreted easily. The coefficient of variation of the RMSE (CVRMSE), that indicates the uncertainty in the model, is the reference metric.
7) Comparative between different regression techniques. The intelligent data analysis techniques taking part in our model-free approach have been applied in scenarios like heating and ventilating systems, prediction of building energy loads, prediction of the outdoor environmental conditions along the time, etc. So far: Bayesian Regularized Neural Network (BRNN) [16], Support Vector Machines (SVM) [17], Gaussian Processes with Radial Basis Function Kernel (GRBF) [18], Random Forest (RF) [19] and Auto-Regressive Integrated Moving Average (ARIMA) [20] are considered. Most of these algorithms are trained to find the best hyper-parameters using a validation strategy.

Finally, the models obtained with the different techniques are evaluated among them with a metric using the test datasets, and the technique with the most accurate result is selected to generate the final predictive model.

4 Examples of data-driven building models

The reference building in which the proposed procedure has been carried out to generate accurate building models is the Technological Transfer Centre (TTC) of the University of Murcia*. This building is used by technological companies and some research groups that collaborate with companies developing industrial scientific projects.

The building has a wide deployment of sensors and devices integrated in the Open Data platform which is working, among other purposes, to improve indoor comfort at the same time that energy is saved.

The first tackled goal is the prediction of energy consumption. For this purpose, we are going to generate the model by selecting the inputs combination and the best data analytic technique that returns the best results. Once we know the most appropriate inputs for our model, we generate the models able to predict them in order to be applied to the energy consumption model in real time.

4.1 Model 1 - Energy consumption prediction

For the prediction of energy consumption we are going to consider the most basic and also common scenario, where the only available data are historical consumption measurements and environmental outdoor observations having as possible inputs: day of the week, month, season, temperature, humidity, radiation, wind speed, dew point, precipitations and density of vapour pressure each hour.

*. www.um.es/otri/?opc=cttfuentealamo

In the cleaning and pre-processing process we deal with noise, outliers and normalize the data by centering (zero mean) and scaling it (standard deviation one). Energy consumption is measured in non regular intervals of time so we are force to come up with a solution.

Taking into account that occupation information is not usually available in buildings - it requires an exhaustive sensor deployment - we have displayed an outline based on basic and logic usability estimations of the building:

- Situation 1 (S1): holidays, weekends and nights (22:00 PM- 06:00 AM)
- Situation 2 (S2): regular mornings (06:00 AM - 14:00 PM)
- Situation 3 (S3): regular afternoons (14:00 PM - 22:00 PM)

Supported by the statistical differences in consumption between situations claimed by a Kruskall-Wallis test ($H(2) = 1307.2$, p-value < 0.01) and the correspondent post-hoc pairwise comparisons using Holm's correction (p-values < 0.01) [21], 3 different models are proposed for each one of the mentioned situations.

- Model S1. Range of energy consumption = [3.578, 14.1] KWh, mean of energy consumption = 7.904 KWh.
- Model S2. Range of energy consumption = [26.01, 86.19] KWh, mean of energy consumption = 54.27 KWh.
- Model S3. Range of energy consumption = [6.357, 53.290] KWh, mean of energy consumption = 31.48 KWh.

The next step is trying to figure out if the considered inputs (outdoor environmental conditions) are related to the output (energy consumption). Except precipitations, all inputs correlate significatively with energy consumption.

Every model gathers the energy consumption during 8 hours, so we have 8 different observations for each environmental input (one each hour). Also, we create two new variables for every attribute by taking its mean and median. Just to clarify the considered inputs, for situation 1 and, for example, temperature, we will have 11 attributes: temperature at 6 AM, at 5 AM, ... at 22 PM, mean of temperature (from 6AM to 22PM) and median of temperature.

Different AI techniques have been used to generate the predictive models of the energy consumption for each one of the situation, these are: GRBF, SVM, MLP, BRNN and RF. Furthermore, several combinations of inputs have been used to train each model separately having that the best results were obtained when using: day of the week, month, season, mean temperature and mean humidity.

After training the models we achieve the best results using the Random Forest (RF) algorithm for situation 1 (mtry = 4, RMSE = 1 KWh) and situation 3 (mtry = 2, RMSE = 3.87 KWh) and Bayesian Regularized Neural Networks (BRNN) for situation 2 (number of neurons = 2, RMSE = 7.08 KWh) representing all these values between a 12.09% and a 12.86% of error (CVRMSE).

Having trained and tested 5 different models, it is necessary to find statistical evidences that the selected one outperforms better not just in a punctual way. Our 10-fold cross-validation with 5 times repetition strategy generates a set of 50 measurements for each model. In Figure 1 it is displayed the statistical performance for situation 3, where red crosses show the median of RMSE for each model. For every situation, the Friedman test, that is the non parametric alternative for repeated measures (within subjects) ANOVA is significant (p-value < 0.05), and looking for corrected pairwise differences, we find that RF is the only one that differs from the others. Between the other models performances there aren't significant differences as can be observed in Figure 2.

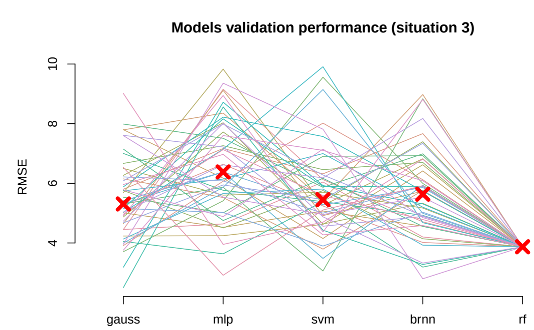

Figure 1: Models validation performance

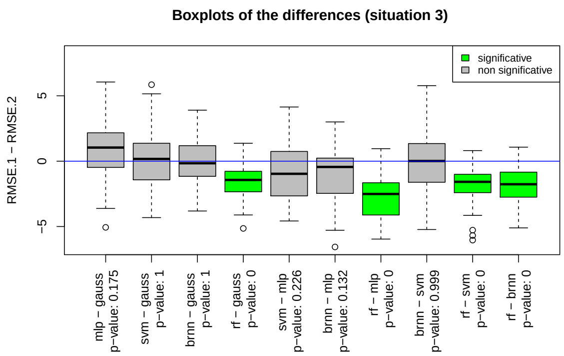

Figure 2: Pairwise differences between models performance

4.2 Model 2 - Environmental conditions prediction

Once the predictive energy consumption models (S1, S2 and S3 models) have been built and evaluated, the next move consists on predicting the inputs of the model in order to apply them in real-time. In our case, the predictive variables are external environmental conditions. There exists several approaches to this problem and we are going to consider time series models such as ARIMA (Autoregressive integrated moving average). They have been widely used in order to predict temperature [20], humidity [22], solar radiation [23], wind speed [24], etc.

We are going to base the pre-processing step on the application of the Box-Cox transformation [25], that stabilizes the variance of the time series and also approximates it to a normal distribution.

In ARMA models the output is expressed as a function of past values or lags (autoregressive part, see Eq. (1)) and past errors or residuals (moving average part, see Eq. (2)).

$$y_t = \mu + \sum_{i=1}^{p} \lambda_i y_{t-i} + \epsilon_t, \tag{1}$$

where μ is a constant, λ_p is the coefficient for the lagged variable in time $t-p$ and ϵ_t is an error term.

$$y_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t, \tag{2}$$

where μ is a constant, ϵ_t is an error term, λ_q is the parameter for the error term in previous periods $t-q$.

With the ARIMA modeling we try to correct violations from the assumption that each element in the series is a random draw from a population with zero mean and constant variance (*white noise*). The deviations that can occur are autoregressive (AR) and moving average (MA), but also, we look for stationarity, which means that mean and variance don't change over time or there are not trends. If the time series show a trend we use the integrated coefficient of the ARIMA. It provides the times that we need to take differences (create new observations by subtracting pairs) in order to eliminate the trend.

ARIMA parameters are:

- p: the number of autoregressive terms,
- d: the number of nonseasonal differences needed for stationarity,
- q: the number of lagged forecast errors in the prediction equation.

Many time series display seasonality, that is, periodic fluctuations. It is logic to think that in general, outdoor conditions present a daily similar fluctuation. The seasonal part of the model consists on terms that are very similar to the non-seasonal components of the model, but they involve backshifts of the seasonal period. The seasonal ARIMA model incorporates both non-seasonal and seasonal factors in a multiplicative model. One shorthand notation for the model is:

$$ARIMA(p,d,q) \times (P,D,Q)_s \tag{3}$$

where the lower case letters refer to the non seasonal part of the model, the capital letters refer to the seasonal part and s is the time span of repeating seasonal pattern.

After setting the model order (p, d, q, P, D, Q), we need to estimate the other parameters involved in equations (1) and (2). The technique used is maximum likelihood estimation (MLE), because it finds the values of the parameters which maximize the probability of obtaining the data that we have observed and so, the logarithm of the probability of the observed data coming from the estimated model is maximized. The usual metric to compare ARIMAS is the corrected Akaike's Information Criterion (AICc) [25].

4.2.1 Strategies of modeling

The collected data always determines the possible strategies of modeling. In our case there are three possible claims.

4.2.1.1 Strategy 1. Historical observations: On the one hand, there exist historical observations of every variable from the year 2013 until nowadays, from the same source as the data with which we trained the models were obtained. This allows us to use univariate time series analysis, where the endogenous variable is going to be explained by its own antique performance. In such scenario, we are losing variability that may be explained by other variables.

The R package forecast [26] permits to use an automation search of the p, d and q parameters with the function *auto.arima()*. In every approach using our data, this function considers that they only present a non-seasonal part, but sometimes it is a fact that it exists. For example, looking at the temperature auto correlation plot (see Figure 3), we see that there exists, at least, a seasonality of frequency 24.

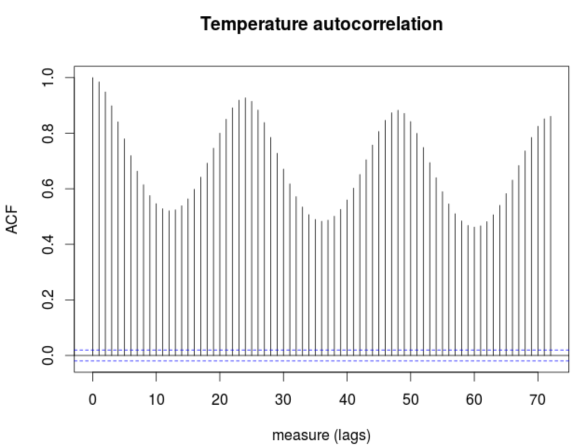

Figure 3: Auto correlation plot for temperature

Bearing this in mind we compare the performance of the automatic process with some ARIMAS whose parameters are setted *a priori* using the function *Arima()*. So, we propose two different performances for this scenario: an automatic one (AA) and a set of already programmed approaches (PA).

4.2.1.2 Strategy 2. External predictions: On the other hand, we have been collecting forecasted data from external open data sources like Weather Underground† every hour with an horizon of 36 hours. This allows us to feed the algorithm not only with past observations, but also using this forecasted variable as regressor. And so, we are going to consider also the Automatic ARIMA with Regressors (AAR) and the Programmed Approaches with Regressors (PAR).

In this case we can also use the machine learning techniques described in section 4.1 - MLP, SVM, BRNN, RF and GAUSS - having as inputs the predictions, and as outputs the observed data.

The biggest problem is that the collection from this source has only been done from December of 2015, having a much more limited sample.

4.2.1.3 Strategy 3. Hybrid strategy: The innovative solution presented in this paper is the combination of the model-free approach described in subsection 4.1 and the ARIMA process above-mentioned. We have 3 candidates to use as inputs in the models:

- External predictions
- Univariate ARIMA predictions
- ARIMA with regressor predictions

In this way, we are combining the predictive power of the above mentioned machine learning techniques with the use of historical data taking into account seasonality as ARIMA does, having for each technique two different models: using as inputs ARIMA predictions with (denoted as $technique_{AR}$) or without regressors (denoted as $technique_{A}$).

4.2.2 Method of evaluation. Rolling windows technique

We need to evaluate the model performance in a time persistent way, having not only a train set and a test set but several ones. In order to asses the model's stability over time we are going to carry out a window rolling analysis of

†. https://www.wunderground.com/

the performance. It can be compared to a leave one out cross validation.

Let set the number of observations of the historical data as n and the prediction horizon as h. This set is initially split into a train set (of length n-h) and a test set (of length h). Then, we train the model and test its performance in the test set computing the RMSE, CVRMSE and R^2. In the second step, we subtract from the historical data as many observations as the selected horizon, having a set of n-h observations. Then, the same method is apply to this subset: splitting into train (length = n-$2h$) and test (length = h) sets, train the model, test it and so on as it is shown in algorithm 1.

Algorithm 1: Rolling window

Require: v: vector of historical data
Given a number of windows w
Given an horizon of prediction h
for $i = 0$ **to** w **do**
v_i = $v[1 : (length(v) - w * h)]$;
l = length(v_i);
train = $v_i[1 : (l - h)]$;
test = $v_i[(l - h + 1) : l]$;
model = $Model$(train);
forecast = $predict$(model, h);
RMSE = $RMSE$(forecast, test) ;
CVRMSE = $\frac{\text{RMSE}}{mean(\text{test})}$;
R2 = R^2(forecast, test) ;
df = $rbind$(df, RMSE, cvRMSE, R2)
end for
return df: data frame with metrics

Traditionally, the rolling window has had a fixed size through the sample but we are sticking to the reality of the application when considering all the previous observations to be part of the train set, as can be observed in Figure 4, where for each step the green block is the test set and the orange block is the train set.

When using the machine learning algorithms we are going to split randomly the train set into a subtrain (75%) and subtest sets (25%). In this way, the best hyper parameters are selected and we predict the next horizon using the model, as we would do in a real-time situation.

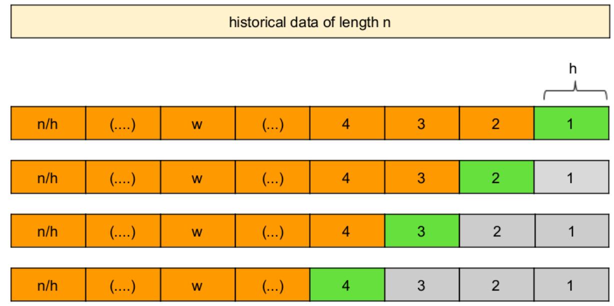

Figure 4: Rolling window schema

4.2 Results

The models are evaluated using the rolling windows strategy for 50 windows. The selected horizon is 24 hours and the outdoor temperature variables that we are predicting are the ones that were selected in Section 4.1 as inputs: temperature and humidity.

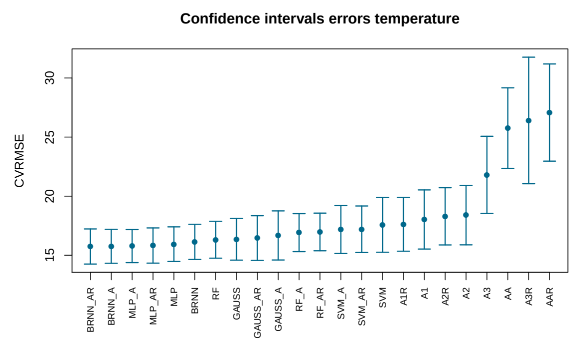

Figure 5: Confidence intervals CVRMSE temperature

We tried 4 different ARIMAS: (1) automatic (AA); (2) Arima(1,0,0)x(0,1,1) (A1); (3) Arima(1,0,1)x(0,1,1) (A2); and, (4) Arima(1,1,1)x(0,1,1) (A3), with and without regressors, having that the better performers are A1 and A1R (i.e. with regressors). Then, for the combination of this method with the machine learning techniques we have used the predictions given by those ARIMAS.

In Figure 5 we appreciate the confidence intervals of the errors (CVRMSE) having that the best one is BRNN combined with AR. It returns a percentage of error with mean 15.79%, and lower and upper confidence intervals of 14.25% and 17.22%. Anyway, it is appreciable that differences between the errors of the first five models are almost indiscernible.

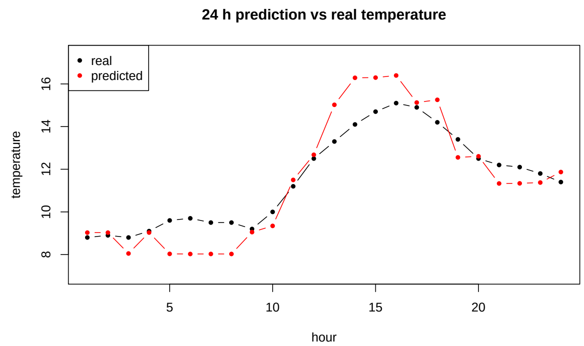

Figure 6: 24 hours predicted vs real temperature

Also, Figure 6 shows one day's prediction using BRNN combined with AR compared with the real observations for temperature.

Doing the same process for humidity we have reached a percentage of error with mean 17.13%, and lower and upper confidence intervals of 14.6% and 19.67%, respectively when using BRNN combined with Arima predictions but these results are closely followed by MLP with Arima, MLP and BRNN.

4.3 Discussion

In this paper we have proposed a methodology to tackle with energy consumption modeling in smart buildings. After collecting data related to energy consumption and outdoor environmental conditions, it is necessary to clean and preprocess these data. Before modeling the prediction of environmental conditions, it is recommended to model energy consumption in order to select the most appropriate inputs of the model and, thus, later predict only the finally selected inputs.

For the energy consumption modeling the steps are: cleaning and preprocessing using normalization - transform attributes in order to have zero mean and standard deviation one, visualization and correlation with the possible inputs. Then, having a first understanding of the data it is possible to propose the study of one or more models, depending on the existence of a partitioning variable - in our case we created 3 situations - supported by an statistical test. Later, we train each model using a validation strategy - in our case, cross validation - in order to find the best hyperparameters of a set of intelligent data analysis algorithms and then, evaluate the performance of each technique with the RMSE metric. Finally, we can also use an statistical test in order to compare the performance - related to this metric - of the different techniques, not just in a punctual way.

For the environmental conditions modeling the steps are similar: cleaning and preprocessing using box-cox transformation, visualization and autocorrelation - in order to determine if there exists seasonality. Then, train the models using different inputs: just the previous observations of the predicted variable - univariate models such as simple ARIMA, some previous observations together with external predictors - ARIMA with regressors and the same intelligent data analysis algorithms than in the previous case - and finally, use the predictions made with the ARIMA modeling together with the predictions coming from external sources. Later, we use a rolling window strategy in order to evaluate the performance of every trained technique which helps us to determine the one (or the ones) that return a smaller error in prediction.

5 Conclusions and Future Work

The main objective of this work is to propose a methodology to generate predictive models of buildings which can be used to implement efficient optimization strategy to save energy. Following this goal, in this paper we focus on modeling energy consumption of buildings associated to the thermal comfort service.

We propose a general procedure to generate the models in charge of predicting the evolution of the main parameters affecting energy consumption. As examples of application of the methodology proposed, we describe the process followed to obtain the predictive model of the energy consumption in a reference building of the University of Murcia, as well as the predictive model of the outdoor environmental conditions involved in the previous model. For providing the final models, different AI techniques have been applied and evaluated according to the results obtained in each case. Results obtained reflect that RF and BRNN are the most suitable techniques to predict energy consumption when inputs like the day of the week, month, season, mean temperature and mean humidity are considered, providing errors between 12.09% and a 12.86%. For the case of the outdoor environmental conditions prediction, the best results for the outdoor temperature prediction are obtained when BRNN is combined with AR to generate the model, which returns a percentage of error with mean 15.79%, and lower and upper confidence intervals of 14.25% and 17.22%. And, following a similar procedure to generate the outdoor humidity prediction, we have reached - with the same techniques combination - a percentage of error with mean 17.13%, and lower and upper confidence intervals of 14.6% and 19.67%, respectively.

The ongoing work is focused on implementing the optimization strategy that, using the models generated in this work, is in charge of optimizing the energy consumption associated to thermal comfort.

Acknowledgments

This work has been partially funded by MINECO TIN2014-52099-R project and ERDF funds, by the European Commission through the H2020-ENTROPY-649849 EU Project, and the Spanish Seneca Foundation by means of the PD program (grant 19782/PD/15).

Acknowledgment

This work has been partially funded by MINECO TIN2014-52099-R project and ERDF funds, by the European Commission through the H2020-ENTROPY-649849 EU Project, and the Spanish Seneca Foundation by means of the PD program (grant 19782/PD/15).

References

[1] D. Petersen, J. Steele, and J. Wilkerson, "Wattbot: a residential electricity monitoring and feedback system," in *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*. ACM, 2009, pp. 2847–2852.

[2] E. Comission, "Directive 2010/31/EU of the European Parlament and of the Council of 19 may 2010 on the energy performance of buildings (recast)," *Official Journal of the European Union*, vol. 53, no. L 153, pp. 13– 34, June 2010.

[3] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Sensing as a service model for smart cities supported by internet of things," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 1, pp. 81–93, 2014.

[4] F. S. Westphal and R. Lamberts, "The use of simplified weather data to estimate thermal loads of non-residential buildings," *Energy and buildings*, vol. 36, no. 8, pp. 847–854, 2004.

[5] A. Rice, S. Hay, and D. Ryder-Cook, "A limited-data model of building energy consumption," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. ACM, 2010, pp. 67–72.

[6] R. Yao and K. Steemers, "A method of formulating energy load profile for domestic buildings in the uk," *Energy and Buildings*, vol. 37, no. 6, pp. 663–671, 2005.

[7] Y. Cheng-Wen and Y. Jian, "Application of ann for the prediction of building energy consumption at different climate zones with hdd and cdd," in *Future Computer and Communication (ICFCC), 2010 2nd International Conference on*, vol. 3. IEEE, 2010, pp. V3–286.

[8] P. A. González and J. M. Zamarreno, "Prediction of hourly energy consumption in buildings based on a feedback artificial neural network," *Energy and Buildings*, vol. 37, no. 6, pp. 595–601, 2005.

[9] S. L. 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] Y. Pan, Z. Huang, and G. Wu, "Calibrated building energy simulation and its application in a high-rise commercial building in shanghai," *Energy and Buildings*, vol. 39, no. 6, pp. 651–657, 2007.

[11] 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.

[12] J. Yang, H. Rivard, and R. Zmeureanu, "On-line building energy prediction using adaptive artificial neural networks," *Energy and buildings*, vol. 37, no. 12, pp. 1250–1259, 2005.

[13] EU Entropy Consortium. (2015-2018) EU Entropy Project. [Online]. Available: http://entropy-project.eu/

[14] M. A. Zamora-Izquierdo, J. Santa, and A. F. Gómez-Skarmeta, "An integral and networked home automation solution for indoor ambient intelligence," *Pervasive Computing, IEEE*, vol. 9, no. 4, pp. 66–77, 2010.

[15] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[16] L. Hawarah, S. Ploix, and M. Jacomino, "User behavior prediction in energy consumption in housing using bayesian networks," in *Artificial Intelligence and Soft Computing*. Springer, 2010, pp. 372–379.

[17] Y. Fu, Z. Li, H. Zhang, and P. Xu, "Using support vector machine to predict next day electricity load of public buildings with sub-metering devices," *Procedia Engineering*, vol. 121, pp. 1016–1022, 2015.

[18] D. J. Leith, M. Heidl, and J. V. Ringwood, "Gaussian process prior models for electrical load forecasting," *Probabilistic Methods Applied to Power Systems*, pp. 112–117, 2004.

[19] H.-x. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012.

[20] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Combining neural networks and arima models for hourly temperature forecast," in *ijcnn*. IEEE, 2000, p. 4414.

[21] J. M. Andy Field and Z. F. Niblett, *Discovering Statistics Using R*, 1st ed. Sage Publications Ltd, 2012.

[22] S. A. Shamsnia, N. Shahidi, A. Liaghat, A. Sarraf, and S. F. Vahdat, "Modeling of weather parameters using stochastic methods (arima model)(case study: Abadeh region, iran)," in *International Conference on Environment and Industrial Innovation. IPCBEE*, vol. 12, 2011.

[23] H. A. Hejase and A. H. Assi, "Time-series regression model for prediction of mean daily global solar radiation in al-ain, uae," *ISRN Renewable Energy*, vol. 2012, 2012.

[24] J. Palomares-Salas, J. De la Rosa, J. Ramiro, J. Melgar, A. Aguera, and A. Moreno, "Arima vs. neural networks for wind speed forecasting," in *Computational Intelligence for Measurement Systems and Applications, 2009. CIMSA'09. IEEE International Conference on*. IEEE, 2009, pp. 129–133.

[25] D. S. S. Robert H. Shumway, *Time Series Analysis and Its Applications With R Examples*, 2nd ed., ser. Springer Texts in Statistics. Springer, 2010.

[26] R. J. Hyndman and Y. Khandakar, "Automatic time series forecasting: the forecast package for R," *Journal of Statistical Software*, vol. 26, no. 3, pp. 1–22, 2008. [Online]. Available: http://ideas.repec.org/a/jss/jstsof/27i03.html