

# Comparing TSK-1 FRBS against SVR for electrical power prediction in buildings

Javier Cózar<sup>1</sup> Gonzalo Vergara<sup>2</sup> José A. Gámez<sup>1</sup> Emilio Soria-Olivas<sup>2</sup>

<sup>1</sup>SIMD Research Group, I<sup>3</sup>A, University of Castilla-La Mancha, Albacete, Spain

<sup>2</sup>IDAL Research Group, E.T.S.E, University of Valencia, Burjassot, Spain

## Abstract

The study of energy efficiency in buildings is an active field of research. Modelling and predicting energy related magnitudes leads to analyse electric power consumption and can achieve economical benefits. In this study, machine learning techniques are applied to predict active power in buildings. The real data acquired corresponds to time, environmental and electrical data of 30 buildings belonging to the University of León (Spain). Firstly, we segmented buildings in terms of their energy consumption using principal component analysis. Afterwards, after test different univariate and multivariate techniques, we applied SVR and a learning FRBS method to compare their performance. Models were studied for different variable selections. Our analysis shows that the FRBS has the lowest error needing a similar learning time than SVR.

**Keywords:** Support Vector Regression, Takagi Sugeno Kang Fuzzy Rule Based System, Power prediction, Buildings electrical power.

## 1. Introduction

Nowadays, building consumption is estimated to be about 40% from total energy consumed in developed countries [1]. The International Energy Agency (IEA) claims that building sector is one of the most interesting sectors to invest in energy efficiency. From the economical point of view, it is possible to reduce the energy consumption in 1509 Mtoe<sup>1</sup> in year 2050. This fact also implies important environmental benefits: if the energy demand is reduced from buildings, the carbon dioxide (CO<sub>2</sub>) emissions could be dramatically reduced too. It can be estimated that 12.6 Gt of CO<sub>2</sub> could not be emitted to the atmosphere in 2050 [2].

In the literature, there is a great variety of data mining techniques that can be used in prediction of energy demanding tasks: linear regression (LR) is applied to predict monthly electrical consumption in large public buildings [3], autoregressive integrated moving average (ARIMA) models let failure detection in electronic equipment [4], and building occupancy is considered to improve predictions [5]. Artificial neural networks (ANN) have been widely

applied to the study of energy consumption in buildings: minimization of the energy to air condition an office-type facility [6], energy of office buildings with daylighting [7] and air condition heating dependence on electrical consumption [8] are some examples. Support vector regression (SVR) is utilized in [9] for short-term electric load forecasting, applying also Fuzzy inductive reasoning from one office to a whole campus electric consumption for one year. A Fuzzy Rule Based System (FRBS) is used in [10] to predict the electricity demand based on temperature changes.

In this study we used real data from 30 different buildings of the University of León (Spain) [11]. These buildings have different uses and can be classified into three types: academic (e.g. faculties), support (e.g. dormitory building, cafeteria, etc.) and research (e.g. technology center).

Our objective is to obtain models with selected techniques and to do one hour prediction. We compare different variable selection for each technique and finally we select the best model. In order to make the predictions we propose the use of a FRBS, due to its high interpretability for the end user. This property is desirable because information about the behaviour of the problem can be understood, and therefore it can be used to improve in some way the performance of the system being monitored (for example, change the time when an activity is carried out in a certain building to avoid a high power consumption at that moment).

In preliminary experiments we have tested different learning techniques to compare with FRBS, using some representative buildings, which are:

**Univariate:** autoregressive (AR), moving average (MA) and a combination of both approaches (ARMA).

**Multivariate:** linear regression (LR), multilayer perceptron (MLP), support vector regression (SVR), regression trees (RT) and model trees (MT).

Univariate methods are not utilised because they do not take into account all the available variables and they are more sensitive to missing values than other methods. Within multivariate methods MLP models are the best in terms of one hour prediction, but the worst in terms of computational time, becoming impracticable for some experiments.

<sup>1</sup>1 toe = 1 tonne oil equivalent = 11.63 MWh

Tree based models are faster than the proposed FRBS method, but errors generated are substantially worse. As we found similar results (for prediction and computational effort) between SVR and FRBS models we decided to select these two techniques for this study.

The rest of the paper is organized as follows. In Section 2 the datasets are detailed. Then, in Section 3, SVR and the proposed FRBS are explained. The experiment setup and results are shown in Section 4. Finally, Section 5 concludes this paper.

## 2. Analysed datasets

From the technical code of spanish buildings [12], the city of León (Spain) is placed in E1 climate zone at 346 m above sea level. This zone has severe climate conditions in winter as well as in summer, hence the electrical consumption due to Heating, Ventilating and Air Conditioning (HVAC) equipment is high (the temperature goes from  $-3.69$  °C in winter to  $38.95$  °C in summer).

The available datasets contain information about environmental, electric and time information variables. The environmental variables were acquired from a single meteorological station, meanwhile, the rest of variables were separately acquired from each building [11]. The variables involved in this study are: hour, day, month, year, ambient temperature (°C), relative humidity (%), solar radiation ( $W/m^2$ ), working day, active power (kW) and standard deviation of hourly active power (kW).

### 2.1. Data processing

The data have been registered each two minutes during 13 months (from first of March 2010 to thirty-first of March 2011), reaching to 285120 instances for each building. In order to obtain an hourly measure per variable, we have calculated the mean for each hour. However, instances having missing values are deleted and only hours containing 30 samples are considered.

To group buildings according to their power behaviour we carried out a principal component analysis (PCA) using a set of additional power related variables: minimum, maximum, mean, range, standard deviation and coefficient of variation of active power for all buildings. Over the obtained PCA space we perform k-medoids clustering. We found that the best choice for the number of clusters, in terms of mean silhouette index, was three (Figure 1). These three groups have 4, 10 and 16 buildings respectively. From a descriptive analysis, we found that the first group of buildings has high hourly power (100 kW) and low power variation (30%), the second group has medium power (30 kW) and high power variation (90%) and the third group has low power (10 kW) and high power variation (100%).

In order to test the two different techniques we have selected three buildings per group. These

buildings are chosen in such a way that they are dispersed all over each cluster (Figure 1). We can see how this selection scans the whole range of first and second principal components which explain the 95.73% of the variation of power related variables. The nine selected buildings with their codes are: **Group 1** (Veterinary (11), Biology (13) and Agriculture (24)), **Group 2** (Technology (5), Philosophy (6) and Sports (17)) and **Group 3** (Cafeteria II (22), Molecular (23) and Dormitory (25)).

Finally, hour variable is transformed to improve the performance of the applied data mining techniques, avoiding the discontinuity between 23 and 0 h. In order to do that, it is common to unfold the hour in two new variables [13]:  $h_x = \sin\left(\frac{2\pi h}{24}\right)$  and  $h_y = \cos\left(\frac{2\pi h}{24}\right)$ , where  $h$  is the hour. In addition, for the sake of simplicity, datasets have been randomly arranged, so we can split the dataset easily for training and testing having data distributed all over the time in both partitions. Notice that as we are making one hour predictions, we do not need time-ordered instances. Finally, inputs for all variables were standardised in SVR method.

## 3. Utilized Techniques

In certain cases, linear techniques, as AR or LR, could be satisfactory [14]. These methods are easy to program and their computational cost is low. When linear techniques are not enough, it is recommended to use non-linear techniques as MLP [15, 16]. MLP have proven robustness and good performance in a broad variety of applications [17, 18] but its main inconvenient is long learning times. Another kind of non-linear and faster successful methods used in regression problems is SVR [16].

On the other hand, Fuzzy Rule Based Systems (FRBSs) has been widely used in regression problems for prediction [19, 20, 21]. Besides being good approximators, the main advantage of this models is the high interpretability for the end user. This can be useful, for example, to understand how our monitored system behaves in order to make changes to improve its performance.

Below, the aforementioned techniques are briefly described.

### 3.1. Support Vector Regression

The basic idea of SVR is to define an interval in which the function that adjusts the dataset do not deviate from the desired value more than a quantity  $\varepsilon$  [16]. Then, it is defined a tube around the desired signal where our predictions must lie in. It is not taken into account the error itself but if this error exceeds  $\varepsilon$ . This approximation allows to treat a regression problem as a classification one, letting to apply the theory of Support Vector Machines (SVM) [16].

It is usual to define the so-called *slack variables*  $\xi_i$  and  $\widehat{\xi}_i$  that indicate when the SVM model fails

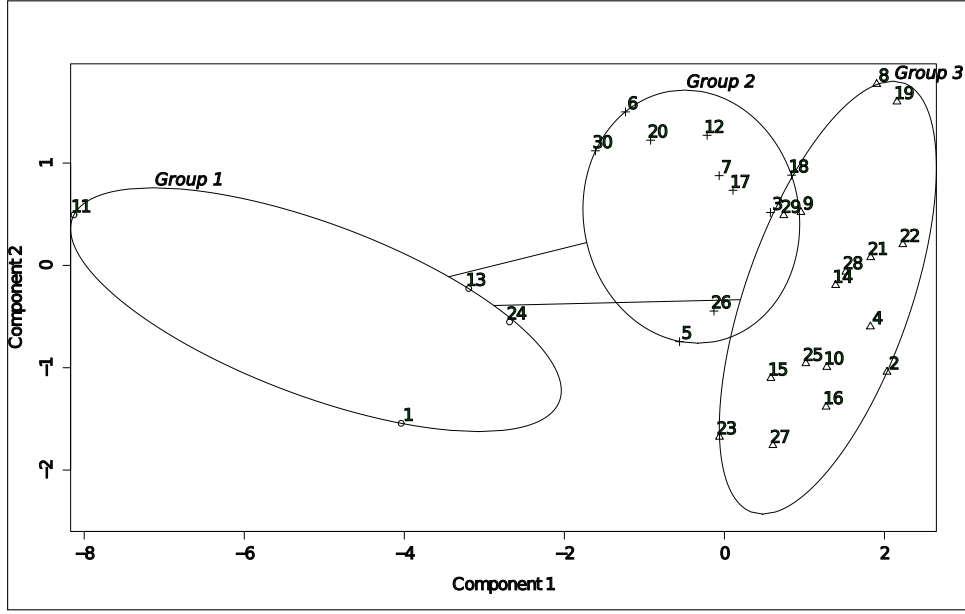


Figure 1: Grouping of the 30 buildings in terms of first and second principal components.

in classification. In our case it means that the output falls outside the tube. Additionally to SVM, it is needed to define a cost function that regularises regression function coefficients. In the linear case, the cost function is given by [16]:

$$\frac{1}{2} \|\mathbf{w}\|^2 + C \sum_i (\xi_i + \hat{\xi}_i)$$

where  $C$  is the penalty factor, and required conditions are:

$$\begin{aligned} y_i - \mathbf{w} \cdot \mathbf{x} - b &\leq \varepsilon + \xi_i \\ \mathbf{w} \cdot \mathbf{x} + b - y_i &\leq \varepsilon + \hat{\xi}_i \\ \xi_i, \hat{\xi}_i &\geq 0 \end{aligned}$$

Then, the learning problem consists in an optimization problem with constraints that can be solved using the classical method of Lagrange multipliers. After equating to zero the partial derivatives of Lagrangian to zero, the function to maximise is [16]:

$$\begin{aligned} &-\frac{1}{2} \sum_{i,j} (\alpha_i - \hat{\alpha}_i)(\alpha_j - \hat{\alpha}_j) \mathbf{x}_i \cdot \mathbf{x}_j \\ &-\varepsilon \sum_i (\alpha_i + \hat{\alpha}_i) + \sum_i y_i (\alpha_i - \hat{\alpha}_i) \end{aligned}$$

where  $\alpha_i$  and  $\hat{\alpha}_i$  are the Lagrange multipliers with conditions  $\sum_i (\alpha_i - \hat{\alpha}_i) = 0$  and  $\alpha_i, \hat{\alpha}_i \in [0, C]$ . In order to get a more flexible model it is usual to introduce the so-called *kernels* [15], [16]. In this case a map from input variables to other space is applied to implement non-linearities in the model. Finally, the error function to be minimized is given by an analogous equation substituting the scalar product  $\mathbf{x}_i \cdot \mathbf{x}_j$  by the kernel function  $k(\mathbf{x}_i, \mathbf{x}_j)$  under the same conditions. A map from input variables to

other space is done to implement non-linearities in the model using functions known as kernels. The kernel function used in this paper is the Gaussian kernel, defined as [16]:

$$k(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}}$$

where  $\sigma$  is the width of the Gaussian function.

### 3.2. Takagi Sugeno Kang order 1 Fuzzy Rule Based Systems

Takagi Sugeno Kang Fuzzy Rule Based Systems (TSK FRBSs) are composed of fuzzy rules with the form:

$$\begin{aligned} R_s : \text{If } X_1 \text{ is } F_1^s \text{ and } \dots \text{ and } X_n \text{ is } F_n^s \text{ then} \\ Y = P_s(X_1, \dots, X_n) \end{aligned}$$

where “ $X_i$  is  $F_i^s$ ” is a predicate,  $X_i$  is the  $i$ -th variable of the problem domain,  $F_i^s$  is a fuzzy set associated to the variable  $X_i$ , and the consequent is a polynomial function of the input variables  $X_1 \dots X_n$ . The order of a TSK FRBS refers to the degree of the polynomial function. Thus, a TSK-1 system means that  $P_s$  is a polynomial of degree 1. In the case of Linguistic FRBSs [22], predicates uses linguistic labels instead of fuzzy sets ( $X_i$  is  $A_i^s$ ), and each linguistic label  $A_i^s$  is related to a fuzzy set  $F_i^s$ . So, in TSK-1 LFRBSs rules are represented as:

$$\begin{aligned} R_s : \text{If } X_1 \text{ is } A_1^s \text{ and } \dots \text{ and } X_n \text{ is } A_n^s \text{ then} \\ Y = b_s + a_{s_1} X_1 + \dots + a_{s_n} X_n \end{aligned}$$

Usually, methods that build automatically a FRBS from data starts from a dataset and a database definition (that is, the variables domains

and the fuzzy sets partition for each variable in the problem). The proposed method [23] learns TSK-1 FRBS models from data in two stages. In the first one, it generates a set of candidate fuzzy rules. After that, it sets the consequent of all candidate rules at once though the well known Least Squares method.

To generate the candidate rules, firstly a fuzzy association rule extraction is carried out, building all fuzzy rules which satisfy that its  $Support(R_s)$  is greater than a given *minimum support*. The support of a rule  $R_s$  is defined as:

$$Support(R_s) = \frac{\sum_{x_p} \mu_A(x_p)}{N}$$

where  $N$  is the number of instances and  $\mu_A(x_p)$  is the coverage degree of the input instance  $x_p$  with respect to the rule  $R_s$ . In this study, we have used the *minimum* as T-Norm to calculate the coverage degree.

However the number of candidate rules can be large, so after that a subgroup discovery procedure is applied to preselect the most interesting rules. To do that, a weighted scheme is applied, in which iteratively the rule with the highest score is selected (and then the score of the remaining rules is recalculated). This iterative process stops when all the instances are covered by at least  $k_t$  candidate rules. In order to obtain the score of each rule, two alternatives can be used: a metric based on its coverage (rewarding the instances which are not covered yet) or a combination of its coverage and the standard deviation for the instances (output variable) that it covers [23].

#### 4. Experimental evaluation

In this section, we carry out a study divided in two steps. First of all, we analyse how useful are the variables of the datasets for each method, that is SVR and FRBS (we need to point out that input and output variables are the same for all the problems). The idea of this first step is to check if some variables can induce noise to the models, or if they simply do not provide useful information to the learning methods (but increase the computational time to run the algorithms).

Once different variable selections has been analysed, we select the best one for each method (the best variable selection can differ from one method to another). After that, we compare SVR versus FRBS, each one executed after applying its corresponding best variable selection.

In order to evaluate each configuration a 10 fold cross-validation has been carried out for the nine buildings selected (three buildings per group). To make a fair comparison, the same 90% of instances for training and the same 10% for test are selected for the configurations which are being evaluated (for

each one of the 10 folds). Cross-validation is a commonly used technique and provide more robust results than others, i.e. holdout train/test split.

The error measure utilised is the *Mean Absolute Error* (MAE) commonly used in electric consumption data (usually with outliers), and more convenient than others measures as the Root Mean Squared Error (RMSE) [24]:

$$MAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N}$$

where  $y_i$  are real data,  $\hat{y}_i$  are predictions and  $N$  is the number of instances. Models are constructed for each fold and mean MAE for each test set is calculated and utilized for the study.

The following parameters have been used for the evaluated techniques. In the case of FRBS, the configuration recommended by the authors has been used [23]: for each domain variable five triangular fuzzy sets are used equally distributed in frequency, the minimum support used for the association rule discovering process is 0.05 and the standard deviation combined with the covering metric has been used to calculate the score of each rule in the pre-screening process. By the other side, several SVR models were trained for each fold scanning parameters  $C$ ,  $\varepsilon$  and  $\gamma = 1/2\sigma^2$  and the best one is selected as the final learnt model. The parameter intervals used are:  $\gamma$  scans 20 equally separated values between  $d-1$  to  $d+1$ , where  $d$  is the mean euclidean distance between all the pairs of points belonging to training set,  $C = \{0.001, 0.01, 0.1, 1, 10, 100\}$  and  $\varepsilon = \{0.01, 0.1, 0.2, 0.3, 0.4, 0.5\}$ .

##### 4.1. Variable selection method

As it was pointed out previously, all the datasets involve the same variables. However, not all these variables are equally useful for the learning methods. In fact, it is possible that some variables lead to the problem of overfitting, or induce noise to the models. Moreover, even if there are some useless variables and taking the whole set of them may not mean a decrease on the performance, the computational time required to build these models are increased indeed. To test if it is worthy to deal with the whole set of input variables, we have classified them into two groups: environmental variables ( $h_x$  and  $h_y$ , day, working day, month, temperature, relative humidity, solar radiation) and power variables (a pair of power and standard deviation for a specific hour  $\{p_t, sd_t\}$ , one  $\{p_{t-1}, sd_{t-1}\}$ , and two hours before  $\{p_{t-2}, sd_{t-2}\}$ ).

First of all, we want to test if environmental variables can be used without the power variables and vice versa. With this approach we can see if power consumption is mostly dependant on the environmental circumstances, or on the contrary this kind of information is irrelevant. In this case, the variable selection consists on the eight environmental

variables ( $\mathbf{E}$ ), and the three combination of power variables:  $p_t$  and  $sd_t$  ( $\mathbf{P}_t$ );  $p_t$ ,  $sd_t$ ,  $p_{t-1}$  and  $sd_{t-1}$  ( $\mathbf{P}_{t-1}$ );  $p_t$ ,  $sd_t$ ,  $p_{t-1}$ ,  $sd_{t-1}$ ,  $p_{t-2}$  and  $sd_{t-2}$  ( $\mathbf{P}_{t-2}$ ). Finally, to test the behaviour of both environmental and power variables together, the combination of the three last cases with  $E$  has been tested:  $E \cup P_t$  ( $\mathbf{EP}_t$ ),  $E \cup P_{t-1}$  ( $\mathbf{EP}_{t-1}$ ) and  $E \cup P_{t-2}$  ( $\mathbf{EP}_{t-2}$ ).

Results (training and test error for each dataset) are shown in Tables 1, 2 and 3 for SVR models and in Tables 4, 5 and 6 for FRBS models. In the case of FRBS, the best variable selection is clearly  $EP_{t-1}$  reaching the best results in 8 out of 9 datasets. In the case of SVR, at first sight  $EP_t$ ,  $EP_{t-1}$  and  $EP_{t-2}$  are the best variable selections, but it is not clear if any of them performs better than the others. Related to the training error, we can see that when we take into account the power variables, the overfitting problem is more noticeable. However, even if this problem is present for  $EP_t$ ,  $EP_{t-1}$  and  $EP_{t-2}$ , these three seem to be the best variable selections for SVR. In the case of FRBS, similar conclusions can be reported, but in that case the variable selection  $EP_t$  seems to behave worse than in the case of SVR.

variable selection	Veterinary		Biology		Agriculture	
	Train	Test	Train	Test	Train	Test
$P_t$	8.808	8.808	6.452	6.468	5.862	5.839
$P_{t-1}$	5.556	6.000	4.223	4.547	5.005	5.186
$P_{t-2}$	5.433	5.830	3.970	4.382	4.792	5.060
$E$	5.561	8.014	3.059	4.704	4.158	6.203
$EP_t$	3.197	<b>4.708</b>	2.510	<b>3.531</b>	3.641	4.269
$EP_{t-1}$	3.256	4.710	2.700	3.552	3.455	<b>4.213</b>
$EP_{t-2}$	3.369	4.792	2.430	3.566	3.109	4.226

Table 1: Training and test MAE of SVR models for buildings in group 1.

variable selection	Technology		Philosophy		Sports	
	Train	Test	Train	Test	Train	Test
$P_t$	3.560	3.539	5.084	5.097	3.544	3.546
$P_{t-1}$	2.426	2.701	2.503	2.875	2.754	2.906
$P_{t-2}$	2.154	2.575	2.369	2.725	2.635	2.889
$E$	2.315	3.626	2.435	3.689	1.861	2.920
$EP_t$	1.302	2.203	1.558	2.156	1.773	<b>2.235</b>
$EP_{t-1}$	1.246	2.188	1.584	2.162	1.733	2.259
$EP_{t-2}$	1.338	<b>2.135</b>	1.510	<b>2.143</b>	1.735	2.284

Table 2: Training and test MAE of SVR models for buildings in group 2.

In order to obtain more information to decide which variable selection is better (specially for the case of SVR), we have ordered them according to its mean rank (Tables 7 and 8). It is also showed the number of wins/ties/losses between the selection with lower mean rank and the rest. We can see that  $EP_{t-2}$  is the best selection according to the rank mean (1.78) in the case of SVR, but it is very close to the other two  $EP_t$  and  $EP_{t-1}$  (both rank means are 2.11). Also,  $EP_{t-1}$  wins 5 out of 9 times against  $EP_t$ , and 6 out of 9 times against  $EP_{t-2}$ .

variable selection	Cafeteria II		Molecular		Dormitory	
	Train	Test	Train	Test	Train	Test
$P_t$	0.265	0.268	2.003	2.004	2.420	2.420
$P_{t-1}$	0.268	0.281	1.804	1.847	2.094	2.153
$P_{t-2}$	0.267	0.280	1.751	1.844	1.928	2.049
$E$	0.381	0.492	1.071	1.805	1.266	2.113
$EP_t$	0.233	0.245	1.285	1.504	1.333	1.687
$EP_{t-1}$	0.225	<b>0.244</b>	1.270	1.500	1.284	<b>1.676</b>
$EP_{t-2}$	0.231	0.248	1.271	<b>1.490</b>	1.349	1.681

Table 3: Training and test MAE of SVR models for buildings in group 3.

variable selection	Veterinary		Biology		Agriculture	
	Train	Test	Train	Test	Train	Test
$P_t$	26.825	26.868	18.829	18.876	17.144	17.194
$P_{t-1}$	5.471	5.796	3.891	4.064	4.671	4.954
$P_{t-2}$	4.688	5.569	3.453	4.035	4.342	4.969
$E$	9.004	9.953	5.321	5.883	7.131	7.770
$EP_t$	5.450	6.361	3.809	4.404	4.627	5.281
$EP_{t-1}$	3.653	<b>4.277</b>	2.730	<b>3.138</b>	3.635	<b>4.100</b>
$EP_{t-2}$	3.562	4.339	2.703	3.196	3.633	4.191

Table 4: Training and test MAE of FRBS models for buildings in group 1.

variable selection	Technology		Philosophy		Sports	
	Train	Test	Train	Test	Train	Test
$P_t$	7.523	7.550	7.540	7.596	9.154	9.204
$P_{t-1}$	2.515	2.628	2.416	2.571	2.443	2.562
$P_{t-2}$	2.190	2.743	2.190	2.656	2.260	2.688
$E$	4.740	5.230	4.669	5.087	3.852	4.198
$EP_t$	2.646	3.074	2.798	3.228	2.917	3.299
$EP_{t-1}$	1.645	<b>1.953</b>	1.794	<b>2.173</b>	1.968	<b>2.274</b>
$EP_{t-2}$	1.619	1.974	1.824	2.200	1.964	2.328

Table 5: Training and test MAE of FRBS models for buildings in group 2.

variable selection	Cafeteria II		Molecular		Dormitory	
	Train	Test	Train	Test	Train	Test
$P_t$	0.373	0.372	5.153	5.171	3.278	3.297
$P_{t-1}$	0.162	<b>0.175</b>	1.696	1.756	1.917	2.031
$P_{t-2}$	0.150	0.280	1.555	1.805	1.733	2.025
$E$	0.452	0.490	2.015	2.200	2.041	2.277
$EP_t$	0.296	0.362	1.592	1.812	1.638	1.875
$EP_{t-1}$	0.176	0.251	1.273	<b>1.440</b>	1.458	<b>1.668</b>
$EP_{t-2}$	0.160	0.255	1.264	1.467	1.448	1.693

Table 6: Training and test MAE of FRBS models for buildings in group 3.

method	rank	win	tie	loss
$EP_{t-1}$ - SVR	<b>1.78</b>			
$EP_t$ - SVR	2.11	4	0	5
$EP_{t-2}$ - SVR	2.11	3	0	6
$P_{t-2}$ - SVR	4.22	0	0	9
$P_{t-1}$ - SVR	5.33	0	0	9
$E$ - SVR	6.00	0	0	9
$P_t$ - SVR	6.44	0	0	9

Table 7: Mean ranks and win/tie/loss metric for SVR models.

method	rank	win	tie	loss
$EP_{t-1}$ - FRBS	<b>1.11</b>			
$EP_{t-2}$ - FRBS	2.11	0	0	9
$P_{t-1}$ - FRBS	3.22	1	0	8
$P_{t-2}$ - FRBS	3.78	0	0	9
$EP_t$ - FRBS	4.78	0	0	9
$E$ - FRBS	6.11	0	0	9
$P_t$ - FRBS	6.89	0	0	9

Table 8: Mean ranks and win/tie/loss metric for FRBS models.

Finally, to give more detailed information about the ranking distributions, we show a boxplot in Figures 2 and 3, where the x-axis refers to the variable selections and the y-axis refers to the mean rank. In this graphics, we can see that in the case of FRBS,  $EP_{t-1}$  is the best selection and the Cafeteria II is an outlier. In the case of SVR, the median of the three best selections equals two: for  $EP_{t-1}$  selection the box is between first and second positions meanwhile for  $EP_t$  and  $EP_{t-2}$  selections their boxes reach the third position. For this reason, and also because it is the best according to the mean rank, we have chosen  $EP_{t-1}$  as the best variable selection method for SVR too.

#### 4.2. Comparison between tested techniques

In this step we put together both SVR and FRBS for the best variable selection  $EP_{t-1}$ . To visualize the results of each method together, we have grouped the test errors for buildings in group 1 in table 9, for buildings in group 2 in table 10 and for buildings in group 3 in table 11.

method	Veterinary	Biology	Agriculture
$EP_{t-1}$ - SVR	4.710	3.552	4.213
$EP_{t-1}$ - FRBS	<b>4.277</b>	<b>3.138</b>	<b>4.100</b>

Table 9: Comparative (MAE for test) between  $EP_{t-1}$  - SVR and  $EP_{t-1}$  - FRBS for buildings in group 1.

method	Technology	Philosophy	Sports
$EP_{t-1}$ - SVR	2.188	<b>2.162</b>	<b>2.259</b>
$EP_{t-1}$ - FRBS	<b>1.953</b>	2.173	2.274

Table 10: Comparative (MAE for test) between  $EP_{t-1}$  - SVR and  $EP_{t-1}$  - FRBS for buildings in group 2.

We can see FRBS wins 6 out of 9 times to SVR. Two of the three times SVR wins is for buildings in the second group, which could mean that for this group SVR is more suitable than FRBS. However, if we take into account the global performance it seems FRBS builds better models than SVR.

method	Cafeteria II	Molecular	Dormitory
$EP_{t-1}$ - SVR	<b>0.244</b>	1.500	1.676
$EP_{t-1}$ - FRBS	0.251	<b>1.440</b>	<b>1.668</b>

Table 11: Comparative (MAE for test) between  $EP_{t-1}$  - SVR and  $EP_{t-1}$  - FRBS for buildings in group 3.

In order to perform a statistical comparison we apply a Wilcoxon Signed Rank test, where the null hypothesis  $H_0$  is that FRBS is equivalent to SVR, and the alternative hypothesis  $H_1$  is that FRBS is better than SVR. The reason for choosing FRBS as the best method in the alternative hypothesis is that we have performed the mean ranks (similarly to the previous subsection), and this method mean rank is 1.33 while it is 1.67 for SVR. Finally, the p-value obtained is  $4.8828 \cdot 10^{-2}$ , so we can say that FRBS is statistically better than SVR with a confidence level of  $\alpha = 0.05$  when the variable selection  $EP_{t-1}$  has been performed.

## 5. Conclusions

In this paper we use data mining techniques to build models to make one hour predictions on the power consumption in different buildings that belongs to the University of León. Each model is built through a learning method, which uses data about a certain building as input. This data contains information about several variables related to that building (temperature, humidity, actual power consumption, etc.).

We want to evaluate the behaviour of a FRBS for this problem, because we consider useful the high degree of interpretability provided by this models, for example to improve the monitored systems in terms of energy power consumption (e.g. rescheduling some tasks to balance the power consumption all over the day).

In order to compare FRBS against other strategy, several data mining techniques has been analysed with a few (due to time limits) of these datasets, and one of them has been selected for the analysis (SVR). The reason beyond this choice is that SVR is the second best method between all the analysed. The first one was MLP, but the time required to perform the analysis was huge, but in the case of SVR the computational time is similar to FRBS.

We have divided this study in two steps. First of all, we have analysed how useful are input variables for each learning method, in order to discard some of them which could induce the models to the problem of overfitting or just increase the computational time without improving the prediction performance.

After that, we have compared the two proposed methods for the best configuration of variable selections. In order to do that, we have performed a statistical Wilcoxon Signed Rank test. Results show

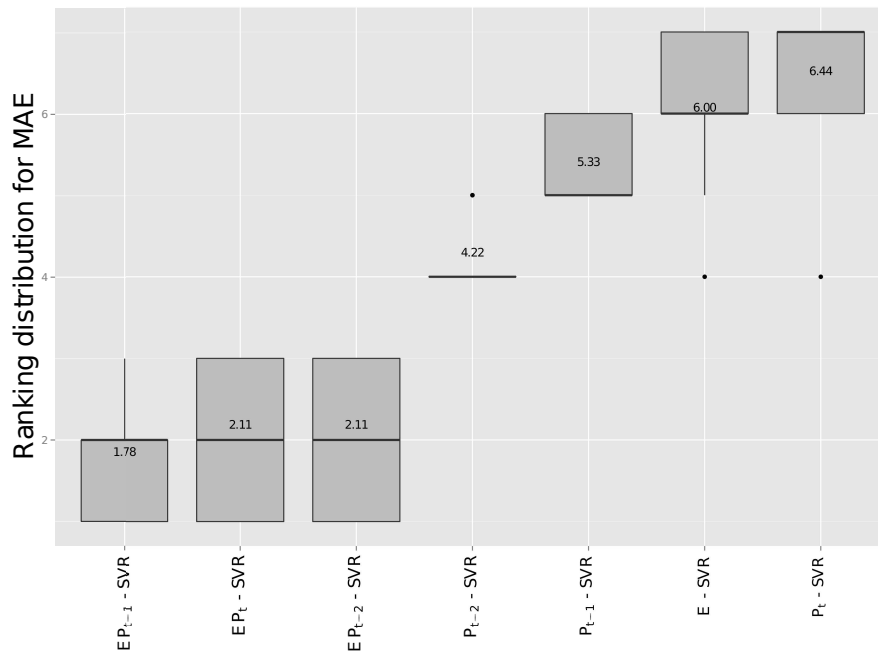


Figure 2: Ranking distribution for SVR.

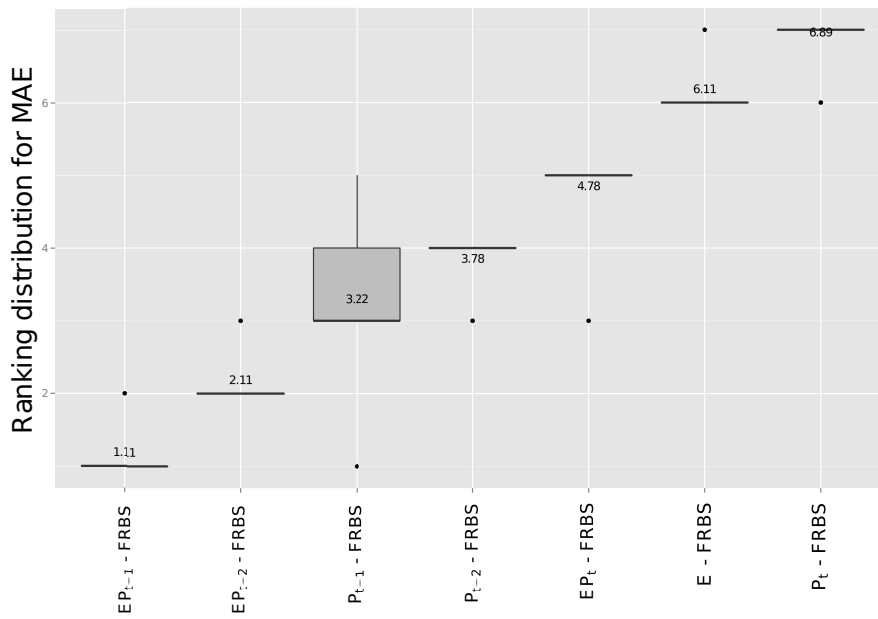


Figure 3: Ranking distribution for FRBS.

that FRBS overcome SVR for the chosen variable selection, with a confidence level of  $\alpha = 0.05$ . Also we can see that the  $E$  configuration, where power is not considered, is the worst configuration with more variables. Finally, the  $P_t$  variable selection is the worst in terms of MAE.

**Acknowledgements.** This work has been partially funded by FEDER funds, the Spanish Government (MINECO) and JCCM through the projects TIN2013-46638-C3-3-P and PEII-2014-049-P. Javier C3zar is also funded by the MECD grant FPU12/05102. Authors want to thank Manuel Dom3nguez of the SUPRESS research group of the University of Le3n for his collaboration.

## References

- [1] U.S. Department of Energy. Buildings Energy Data Book, 2010.
- [2] I.E.A. International Energy Agency. Energy Performance Certification of Buildings, 2013.
- [3] Y. Ma, J.Q. Yu, C.Y. Yang, and L. Wang. Study on power energy consumption model for large-scale public building. In *Proceedings of the 2nd international workshop on intelligent systems and applications*, pages 1–4, 2010.
- [4] X. Bian, Q. Xu, B. Li, and L. Xu. Equipment fault forecasting based on a two-level hierarchical model. In *Automation and Logistics, 2007 IEEE International Conference on*, pages 2095–2099, 2007.
- [5] G.R. Newsham and B.J. Birt. Building-level occupancy data to improve arima-based electricity use forecasts. In *Proceedings of the 2nd ACM workshop on embedded sensing systems for energy-efficiency in building*, BuildSys '10, pages 13–18. ACM, 2010.
- [6] A. Kusiak, M. Li, and F. Tang. Modeling and optimization of HVAC energy consumption. *Applied Energy*, 87(10):3092–3102, 2010.
- [7] S.L. Wong, K.K.W. Wan, and T.N.T. Lam. Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*, 87(2):551–557, 2010.
- [8] B.B. Ekici and U.T. Aksoy. Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, 40(5):356–362, 2009.
- [9] S. Jurado, J. Peralta, A. Nebot, F. Mugica, and P. Cortez. Short-term electric load forecasting using computational intelligence methods. In *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, pages 1–8. IEEE, 2013.
- [10] H. Shakouri G., R. Nadimi, and F. Ghaderi. A hybrid tsk-fr model to study short-term variations of the electricity demand versus the temperature changes. *Expert Systems with Applications*, 36(2):1765–1772, 2009.
- [11] S. Alonso. *Supervisi3n de la energ3a el3ctrica en edificios p3blicos de uso docente basada en t3cnicas de miner3a de datos visual*. PhD thesis, Departamento de Ingenier3a El3ctrica, Electr3nica, de Computadores y Sistemas. Universidad de Oviedo., 2012.
- [12] Ministerio de Fomento, Gobierno de Espa3a. C3digo T3cnico de la Edificaci3n, 2010.
- [13] O.A.S. Carpinteiro, A.P. Alves da Silva, and C.H.L. Feichas. A hierarchical neural model in short-term load forecasting. In *IJCNN (6)*, pages 241–248, 2000.
- [14] S.A.H. Soliman and A.M. Al-Kandari. Electric load modeling for long-term forecasting. In *Electrical Load Forecasting*, pages 353–406. Butterworth-Heinemann, Boston, 2010.
- [15] S. Haykin. *Neural Networks and Learning Machines*. Prentice Hall, 3 edition, 2009.
- [16] C.M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., 2006.
- [17] M. Paliwal and U.A. Kumar. Neural networks and statistical techniques: A review of applications. *Expert Systems with Applications*, 36(1):2–17, 2009.
- [18] A. Vellido, P.J.G Lisboa, and J. Vaughan. Neural networks in business: a survey of applications (1992–1998). *Expert Systems with Applications*, 17(1):51–70, 1999.
- [19] L.X. Wang and J.M. Mendel. Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems and Man and Cybernetics*, 22(6):1414–1427, 1992.
- [20] O. Cord3n and Herrera F. A two-stage evolutionary process for designing tsk fuzzy rule-based systems. *IEEE Transactions on Systems and Man and Cybernetics and Part B: Cybernetics*, 29(6):703–715, 1999.
- [21] R. Alcal3, J. Alcala-Fdez, J. Casillas, O. Cord3n, and F. Herrera. Local identification of prototypes for genetic learning of accurate tsk fuzzy rule-based systems. *International Journal of Intelligent Systems*, 22(9):909–941, 2007.
- [22] L. A. Zadeh. The concept of a linguistic variable and its application to approximate reasoning-i. *Information sciences*, 8(3):199–249, 1975.
- [23] J. C3zar, L. de la Ossa, and J.A. G3mez. Generaci3n de reglas difusas tipo tsk-1 basada en el principio apriori derivando el sistema de reglas mediante b3squeda local. In *Actas del X Congreso Espa3ol sobre Metaheur3sticas, Algoritmos Evolutivos y Bioinspirados (MAEB15)*, 2015.
- [24] C.J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate Research*, 30(1):79, 2005.