Contents lists available at
ScienceDirect

Journal of Building Engineering
journal homepage: www.elsevier.com/locate/

A case study on understanding energy consumption through prediction and visualisation (VIMOEN)

L. G. B. Ruiz[a,*], M.C. Pegalajar[a], M. Molina-Solana[b,a] and Yi-Ke Guo[b]

bacaruiz@ugr.es, mcarmen@decsai.ugr.es, m.molina-solana@imperial.ac.uk, y.guo@imperial.ac.uk

[a] *Department of Computer Science and Artificial Intelligence, University of Granada, Spain.*
[b] *Data Science Institute, Imperial College London, UK*

Abstract

Energy efficiency has emerged as an overarching concern due to the high pollution and cost associated with operating heating, ventilation and air-conditioning systems in buildings, which are an essential part of our day to day life. Besides, energy monitoring becomes one of the most important research topics nowadays as it enables us the possibility of understanding the consumption of the facilities. This, along with energy forecasting, represents a very decisive task for energy efficiency. The goal of this study is divided into two parts. First to provide a methodology to predict energy usage every hour. To do so, several Machine Learning technologies were analysed: Trees, Support Vector Machines and Neural Networks. Besides, as the University of Granada lacks a tool to properly monitoring those data, a second aim is to propose an intelligent system to visualize and to use those models in order to predict energy consumption in real-time. To this end, we designed VIMOEN (VIsual MOnitoring of ENergy), a web-based application to provide not only visual information about the energy consumption of a set of geographically-distributed buildings but also expected expenditures in the near future. The system has been designed to be easy-to-use and intuitive for non-expert users. Our system was validated on data coming from buildings of the UGR and the experiments show that the Elman Neural Networks proved to be the most accurate and stable model and since the 5th hour the results maintain accuracy.

Keywords: energy efficiency; energy forecasting; visualization; Mapbox; energy monitoring

* Corresponding author. Tel.: +34 958-241-000 / Ext. 20466; e-mail: bacaruiz@ugr.es. At c/ Periodista Daniel Saucedo Aranda, s.n., 18071 Granada, Spain.

1. Introduction

Buildings are known to be among the largest consumers of energy at the global level (International Energy Agency, 2013). In the USA (Pérez-Lombard, Ortiz & Pout, 2008) HVAC systems account for 20% of the global consumption, and the US Department of Energy calculates that buildings are responsible for 70% of the electricity in the country; the amount of energy consumed in the European Union reaches more than 50%, and other countries such as China or Iran have increased their consumption in more than 10% in the past 20 years (Chen, Shi, Shen, Huang & Wu, 2019). A considerable portion of this energy consumption in buildings is linked to heating, ventilation and air conditioning systems —HVAC— which oversee maintaining comfort for the building's occupants. Typically, these HVAC systems run on rigid schedules and preloaded rules and do not use any external and dynamic information in order to optimize energy efficiency (Gelazanskas & Gamage, 2014; Martínez-Municio, Rodríguez-Benítez, Castillo-Herrera, Giralt-Muiña & Jiménez-Linares, 2018). Besides, the growth rate of population and the necessity of energy conservation in this sector have become a significant concern for many governments in the world. Increasing levels of energy consumption are correlated not only with the price of the energy of each country but also with environmental pollution and its negative impact on health. Recent studies (García-Hinde, et al., 2018; Santamouris & Kolokotsa, 2015) maintain that global climate change has seriously increased the frequency of extreme weather conditions, bringing with it severe energy impact, high temperature, extreme climatic conditions and pollution affecting the environmental comfort conditions and, what is more important, health. Hence, the need for prompt action to prevent the imminent impact of climatic change and aggravating overheating on the energy consumption in buildings discloses the importance and seriousness of the problem.

In order to achieve more efficient buildings (Martínez-Municio et al., 2018), it is necessary to understand how buildings consume energy (Lopes, Antunes & Martins, 2012) and have suitable monitoring systems to reduce energy waste and generate energy savings. Granderson and Lin (Granderson & Lin, 2016) indeed pointed out that building energy information systems are a powerful technology to monitor and analyse consumption which leads to considerable energy and monetary savings. Nevertheless, appropriately handling the vast amount of available data is not a trivial undertaking because of heterogeneity in data provided by the sensors and smart devices (Molina-Solana, Ros, Ruiz, Gómez-Romero & Martin-Bautista, 2017).

In fact, several building energy management systems have been developed as analysis software for data acquisition and monitoring in diverse scenes. For instance, Sarma et al. (Sarma, Singh & Bezboruah, 2018) presented a design of a low-cost data acquisition system for data monitoring, data storing and plotting of live streaming data; Tae-Keun et al. (Oh, Lee, Park, Cha & Park, 2018) proposed a three-dimensional visualization solution for building energy management targeting recommendations to enhance the energy efficiency obtained from the energy consumption; and Pahl et al. (Pahl, Goodhew, Boomsma & Sheppard, 2016) aimed at energy visualization, in which they investigate how to reduce energy consumption in buildings by combining psychology principles and intelligent techniques. As mentioned above, there are plenty of researchers and different approaches to address this problem. Another interesting research line different from those already cited is the use of language modelling (Natconsumers, 2015). Thanks to linguistic summaries it is possible to provide a description of the consumption in natural language (Martínez-Municio et al., 2018), or even to prototype the energy time series to extract hidden information from the data (Leon-Alcaide, Rodriguez-Benitez, Castillo-Herrera, Moreno-Garcia & Jimenez-Linares, 2020; Moreno-Garcia, Rodriguez-Benitez, Jimenez-Linares & Triviño, 2019).Many authors elaborate designs based on graphs, tables and figures whose purpose is to depict data in a more organized manner, but rare are the works that propose interactive and visual tools for this purpose. A notable exception is (Murshed, Al-Hyari, Wendel & Ansart, 2018) where a design of a web application for smart city visualization was implemented. Furthermore, in almost all scientific studies the solution is focused on just one building, not on a whole set of facilities distributed in diverse areas.

In addition to those, there are many solutions in literature for improving energy efficiency and many intelligent algorithms for processing and managing the energy consumption data, indeed, there are some studies whose goal is the cut of energy consumption to reduce buildings footprint, for a complete review of these solutions please refer to (Williams, et al., 2016). An example of this is the work of Luzi et al (Luzi, Vaccarini & Lemma, 2019) who propose a methodology to predict indoor and outdoor thermal loads, and by minimizing the temperature alteration enables efficient use of energy. In this case, the visualization of the results is merely offered for informative purposes. Another example of this can be found in (Ahn & Cho, 2017) in which Fuzzy Inference Systems and Artificial Neural Networks are compared to analyse the energy efficiency and control benefits. In (Shiel & West, 2016) the effects of external variables in buildings operating systems were studied to enhance the accuracy of the models and they show their results in the classical regression graphs. In the same vein, in a previous work of the same authors (Shiel & West, 2015), the research is focused on analyse their results by providing a huge table with the outputs obtained by the whole process in which the task of drawing any conclusion is not easy at first sight because of the high information depicted. An interesting recent study is carried out in (Runge, Zmeureanu & Le Cam, 2020) that proposes an hybrid short-term predictive solution to model electric demand, unlike the rest of the works mentioned so far, in this work the authors employ a carpet plot that can be very illustrative and useful to extract information, however, in this case, they also choose a non-dynamic solution. Probabilistic Algorithms and Model predictive control we used in (Gómez-Romero, et al., 2019) reporting energy savings around 35% in the intermediate winter season in an office building in Helsinki.

We can find many other examples like these in which the visualization piece of the solutions has a static informative role or simply does not exist (Aznavi, Fajri, Sabzehgar & Asrari, 2020; Bilous, Deshko & Sukhodub, 2018; Tian, Si, Shi & Fang, 2019). However, if one tries to find visualization solutions, then they lack a sophisticated intelligent model that helps in the decision-making (Elbeltagi, Wefki, Abdrabou, Dawood & Ramzy, 2017; Zhu, Saeidi, Rizzuto, Roetzel & Kooima, 2018). One of the closest works to those which address both solutions can be found in (Najjar, Figueiredo, Palumbo & Haddad, 2017). The latter integrates a building information modelling system that delivers distinctive intelligent modelling and information management but based on the data currently available and does not provide any prediction or integrate any artificial intelligent model for that purpose.

As we were saying, in nearly all instances, data are employed without the need for being visualized, using consumption data only for this purpose. The main problem here arises when non-expert users attempt to use those technologies. The task becomes an arduous and laborious issue, due to the learning difficulties to understand panels with too much information, and the relevant information seems to have less importance; making it very difficult to control and manage the energy. Thus, the basic requirement is the need to delivering intuitive and easy-to-use software to the users and operators in order to ensure they understand what happens to the energy of their buildings. It should be noted that, using a crowded dashboard of information, or using algorithms and intelligent systems only are non-perfect solutions because of its limited interpretability. The proposal we describe in this work precisely pursues this objective to improve current solutions by simplifying and clarifying the knowledge to assist users in making faster diagnoses, which can improve decision making and efficiency.

Hitherto, very little research has been done in energy efficiency and visualization incorporating new intelligent technologies. Most recent studies are focused on applying data mining techniques (Deb, Zhang, Yang, Lee & Shah, 2017; Ruiz, Cuellar, Delgado & Pegalajar, 2016; Ruiz, Rueda, Cuéllar & Pegalajar, 2018) on the one hand, and data visualization (Huacón

& Pelegrin, 2017; Rodrigues, et al., 2017) on the other, without combining both approaches. Thus, to achieve adequate monitoring of the consumption in a building complex and adding intelligent techniques for effectively realizing energy savings, this work presents a web-based application for monitoring the energy consumption of distributed buildings following a more visual strategy and thus more user-oriented.

Information visualisation has been extensively demonstrated as a viable tool for improved data understanding, as it increases concentration and makes information more attractive, hence supporting more informed decision-making (Oh et al., 2018). For this reason, we have developed an innovative application using recent web technologies and energy consumption forecasting methods to predict the waste of energy and improve the energy efficiency in a set of buildings located in different geographic zones. We adopted the Mapbox GL JS API (MapBox & LLC, 2018) to render an interactive map with buildings' locations and we used several Machine Learning models to predict consumption.

Due to the high difficulty of accurately forecasting energy consumption, diverse approaches can be found in the literature to solve this problem. Some of those models have recently gained considerable popularity: linear regression (Fumo & Rafe Biswas, 2015), regression trees (Nagy, Barta, Kazi, Borbély & Simon, 2016), autoregressive moving average and grey models (Yuan, Liu & Fang, 2016), fuzzy-based models (Davlea & Teodorescu, 2016), support vector machine (Jung, Kim & Heo, 2015), neural network (Leon-Alcaide et al., 2020) among many other hybridizations. Historical data of the energy consumption of the University of Granada (UGR) will be used in the comparative study. Most of the buildings present more than five years of data. This data was acquired from sensors installed in buildings.

Our first goal is to compare those techniques so as to obtain the most accurate approach for bringing it into the visualization system.

Recent publications have revealed increased interest in energy visualization due to his huge potential regarding the challenges surrounding energy efficiency, global warming and exploitation of energy resources. In addition to this, it is important to mention that the UGR currently lacks a support tool for monitoring or access adequately to energy consumption data. Consequently, to fill this gap and solve this issue, our second target in this work is to provide a unique software which combines soft computing and visualization techniques to create a monitoring system endowed with intelligence. For that purpose, we focus our efforts on design an application to centrally monitor a set of distributed buildings. In doing so, we procure a visual framework to manage the current state of the buildings and to know at a glance unexpected changes in their behaviours, to eventually identify possible causes of low performance by using forecasting models.

This tool (called VIMOEN) has been tested with data from buildings of the University of Granada (UGR). 25 buildings have been considered, in 7 campuses located in southern Europe and North Africa. VIMOEN's visualization platform will enable accurate energy consumption management and will allow anticipating future scenarios in the near future as well as providing updated reports about buildings.

In summary, the overall aim of this work is to design and implement an energy consumption visualization tool. This software will provide energy information about the energy information on the past, present and future state of a distributed group of buildings spread geographically throughout diverse areas. The system has been designed to ease the use and to improve the interpretability of the energy information collected from buildings in order to aid any user who needs to know in real-time how their buildings are consuming, avoiding overstuffed dashboards which often result complex and difficult to interpret. Additionally, adding to it intelligent forecasting models to anticipate and manage consumption changes and to make effective planning and decision making about energy expenditure.

The rest of the paper is structured as follows: Section 2 introduces the forecasting techniques for energy consumption. Section 3 describes the data used in this study. Section 4 presents the methodology carried out, data pre-processing and VIMOEN, along with its different parts. Section 5 gathers all outcomes related to energy forecasting and demonstrates the usage of VIMOEN. The last section concludes with some comments and proposals for future works.

2. Forecasting techniques

This section contains a short presentation of the three different types of machine learning techniques used to predict energy consumption: trees, support vector machine models and artificial neural networks.

2.1. Regression Tree

Regression Tree (RT) is one of the simplest models among all the machine learning techniques. Tree-based models are commonly used in both classification and regression problems. Although, in the first case, it is called Decision Tree. Its tree-structured form is based on binary recursive partitioning of a dataset. This process divides observations into different groups and is utilized to decide the structure of the tree (Xu, Watanachaturaporn, Varshney & Arora, 2005). The main advantage of RT over classical statistical methods is that RT does not make a priori assumptions concerning the dependencies between input and output variables. Furthermore, thanks to its configuration RT allows the modelling of nonlinearities in data. CART is the standard nonparametric procedure to perform the predictor selection. In this study three varieties of RT were used: the simple RT, Boosted Regression Tree (BoRT) and Bagged Regression Tree (BaRT) (Prasad, Iverson & Liaw, 2006). BoRT and BaRT are an ensemble (Ribeiro, Mariani & Coelho, 2019) approach which combines the results of several RT to predict the target output. The main difference between those two approaches is that, while BaRT creates an aggregated model with less variance, BoRT focus new predictors on observations that others get wrong.

2.2. Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning method which was created to classify a dataset into two separable classes. To do so, SVM finds the best hyperplane by maximizing the distance between the two classes. However, SVM has recently gained popularity to solve regression problems and have shown to be superior to traditional empirical risk minimization approaches. SVM turns the input space into a high-dimension space and models data according to a kernel function (Jung et al., 2015). We implemented two versions of the SVM regression model for this work: the first is the Linear SVM (L-SVM) model which maps data into a linear function. This model presents high interpretability and low flexibility. In contrast, the second model is the Coarse Gaussian SVM (CG-SVM) whose kernel is the radial basis function and kernel scale is set to $4 \cdot \sqrt{P}$, being P the number of predictors. This model has a hard interpretability and a rigid response function. The main advantage of SVMs is finding the best and unique solution to minimize the objective function.

2.3. Artificial Neural Networks

Lastly, Artificial Neural Networks (ANNs) are a particular machine learning technique inspired by brain connections. ANNs are non-mathematical parametric models capable of modelling non-linear dependencies between input and output variables. In addition to the previous models, ANNs are widely used in both classification (Duque-Pintor, Fernández-Gómez, Troncoso & Martínez-Álvarez, 2016) and regression (Iruela, Ruiz, Pegalajar & Capel, 2020; Ruiz, 2020) problems. Due to its positive results in numerous problems, ANNs have gained great popularity as of lately, with many types of ANNs found in the literature. The simplest one is the Multi-Layer Perceptron (MLP) which we used in this study to predict energy consumption (Ruiz, Capel & Pegalajar, 2019). Also, as we needed the modelling of time dependencies of data to predict future consumption usage, the Elman Neural Network (ENN) has been utilized as well, as this introduces the concept of *memory*. To this end, ENN uses positive feedback to fit this memory in the model. This feature provides a better modelling of time-series with historical dependencies (Fu, 2011).

3. Dataset

As a case study to test the feasibility of the forecasting models and VIMOEN, we worked with data collected from University of Granada (UGR)[1]. UGR is located and distributed in three different cities: Granada (south of Spain, Europe), Ceuta and Melilla (north of Africa). UGR is divided into 7 campuses (*Centro, Fuentenueva, Cartuja, Ciencias de la Salud, Aynadamar, Ceuta* and *Melilla*) each one of them comprising different types of buildings, and consequently, different sensors technologies and energy consumptions need to be considered and dealt with; e.g., teaching units, faculties, schools, departments and training centres. An additional difficulty arises from the disparity of reporting software and different information registers, as each building was commissioned independently. Luckily, most of them register energy consumption and temperature.

As noted above, UGR has recently installed technologies to support building monitoring; and those pieces of software are designed to collect and transform energy data from all sensors and meters in a usable form. The data is stored in a database which gathers monitored information such as consumption, energy and power, as well as other external data such as temperature and humidity, which do not come from the installed sensors but from third party's services. This smart management system captures all this information in real-time and is used by the University for monitoring and analysing buildings' consumption. However, the building automation system is in charge of storing raw data as it comes from sensors. This makes it difficult to work with such information. For this reason, the treatment of those data is essential and must be preprocessed. Fig. 1a illustrates an example of the data from meters. As one may see, it is very hard to obtain any information as the consumption is stored considering how much energy has been used up to now. This is why, in Fig. 1a, a rising trend is depicted, until it reaches breaking point where the meter resets its counter. However, once raw consumption data are processed,

[1] Due to Data Protection Laws and security issues, we are unable to share the actual data of energy consumption; this fact however does not affect the development and results of this research.

they present the form shown in Fig. 1b which details a more usable way of the same information, and a particular behavior may be appreciated. Besides, this treatment is vital due to possible problems in data, such as, problems during the data transfer, fault detection, broken devices or system failure.

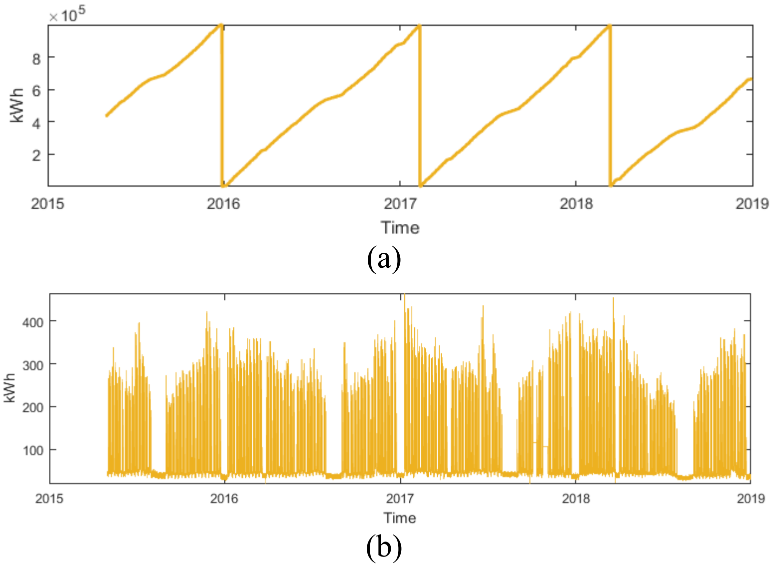


Fig. 1. Example of (a) the raw data extracted from a building's meter (b) and the transformed and processed energy consumption in an hourly time scale. Note the different scales of magnitude.

4. Methodology

This study has two main goals or stages. The first one is to find the best forecasting model in order to predict the energy consumption in buildings. The second one is to utilize past, present and future building's information so as to help improve their consumption monitoring.

4.1. Data-related treatment

In order to take up the first problem, a retrieving data procedure was designed to extract hourly energy consumption from buildings. As mentioned above, data present several problems and its form when stored by meters is not usable or readable directly (see Fig. 1a). Data were preprocessed through a noise treatment and data transformation. The first procedure was carried out with two solutions: 1) Moving-average filter to remove outliers (Kayacan, Ulutas & Kaynak, 2010); 2) linear interpolation using the neighboring grid values of the consumption to fill missing data (Andiojaya & Demirhan, 2019). The data extracted from the electric meters were stored as a historical time-series which represents the accumulated amount of energy used to date. In other words, one record of the raw data denotes the aggregate sum of all preceding consumption until at that time. Hence, the second operation involves changing the raw data into an hourly time scale. In this way, one sample of the data is now the consumption spent that hour.

Once that was done (see Fig. 1b), we can concentrate on the forecasting goalby testing several machine learning algorithms: RT, SVM and ANN, along with some variants of these models. We work with historical data, and as a result, a time-series y corresponding to a particular building will be modelled following the equation:

$$y(t+1) = m\big(y(t), y(t-1), y(t-2), \dots, y(t-p)\big) + \epsilon(t+1) \tag{1}$$

In this way, a future value $t+1$ depends on the p past values, and the specific model m will adjust those data. The error of the forecasting model at time $t+1$ is represented by $\epsilon(t+1)$.

Data are also normalized in range $[0, 1]$ in order to guarantee that the models do not give more weight to those attributes with higher range in consumption by using the next equation:

$$y = \frac{y - y_{\min}}{y_{max} - y_{min}} \tag{2}$$

Besides, we applied Root Mean Squared Error (RMSE) to check the accuracy of the models:

$$RMSE = \sqrt{\frac{1}{N} \cdot \sum_{t=1}^{N} \big(D(t) - O(t)\big)^2} \tag{3}$$

Where $D(t)$ is the desired response from the forecasting model, $O(t)$ is the output at instant t and N is the number of instances.

Fig. 2 illustrates the process flow to train the predictive models. As a summary, this figure gathers all the operations that are carried out. Firstly, the raw data are collected from diverse sensors installed in the buildings. This information is pre-processed in order to allow us to work with the given data in a useful way. In this stage, data granularity is set according to the established time scale, in this case, hourly. Once data are transformed into the desired form, they should pass the missing and noising values procedures previously mentioned. Thus, the wrong data and gaps in consumption are treated. As soon as the data are considered clean, they are split into two sets: training and test. The first one is employed to train the model, and the second one is used to validate and confirm that the forecasting model has learnt satisfactorily to predict consumption. Once the model is considered validated, it is stored to later predict the energy consumption when required.

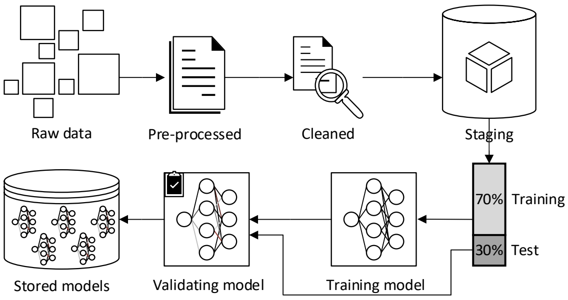


Fig. 2. Flowchart of the training process of forecasting models. This follows the usual procedures and best-practices in Machine Learning.

4.2. Proposed visualization system

VIMOEN is composed of three main components: database, forecasting models and visualization module. The system has been developed using the Mapbox GL JS library to render interactive maps, ANNs to predict the consumption of the buildings and data from the facilities of the UGR to create the design of the system as illustrated in Fig. 3. The first part represents the origin of the data. All information is extracted from the buildings through the automation system which collects the data from the physical world by using the sensing technologies. Each sensor provides different information about the buildings. Contextual information of the buildings must be pre-processed and transformed before it can be used to provide useful knowledge.

Consequently, a next step is to join and process such information in order to properly handle and manipulate it. Thus, the database gathers the remodelled data of the hourly energy consumption or the temperature, in addition to other metadata about the building such as its geographic location.. All that knowledge will be used to build the final system, but firstly data on energy consumption is employed by the predictive models to forecast forthcoming consumption.

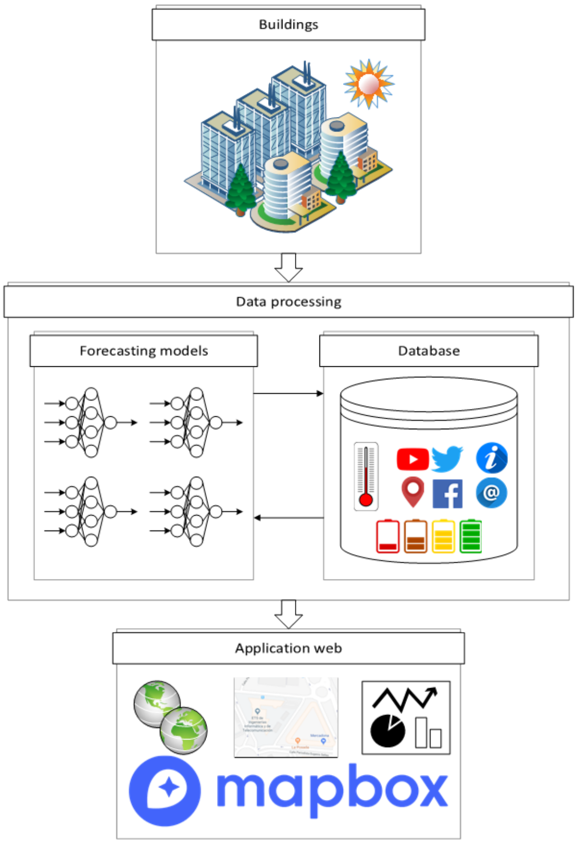


Fig. 3. Overall scheme of VIMOEN.

Our visualisation tool, VIMOEN, has been designed to manipulate the information of the buildings and to depict it in the form of knowledge. And thus, to clearly visualize the energy consumption of the whole distributed institution in a centralized system. This component collects the buildings' data and links them for visualization in client-side by using web browser. It also incorporates the forecasting models already trained to include the predicted consumption of the buildings. For visualization purposes, the Mapbox's JavaScript library is used here to render the maps. Geographical data about the location of UGR's buildings was available for our test.

An illustration of the proposal layout for the energy visualization can be found in Fig. 4. The first layer is provided by the Mapbox API with the street layer and all the system is based on it. The map layer includes methods and properties which allow change the appearance of the map and enable users to interact with it. The second layer includes the location of all buildings. The buildings' geometry is determined with a two-value coordinates, the longitude and the latitude, both measured in degrees. Each point is endowed with a popup component in which the real-time consumption is depicted. Furthermore, each point offers useful information at a glance and provides easy access to historical consumption (as described in the next section).

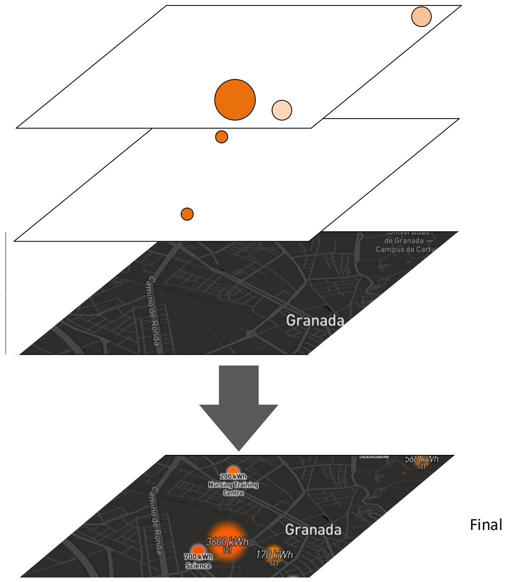
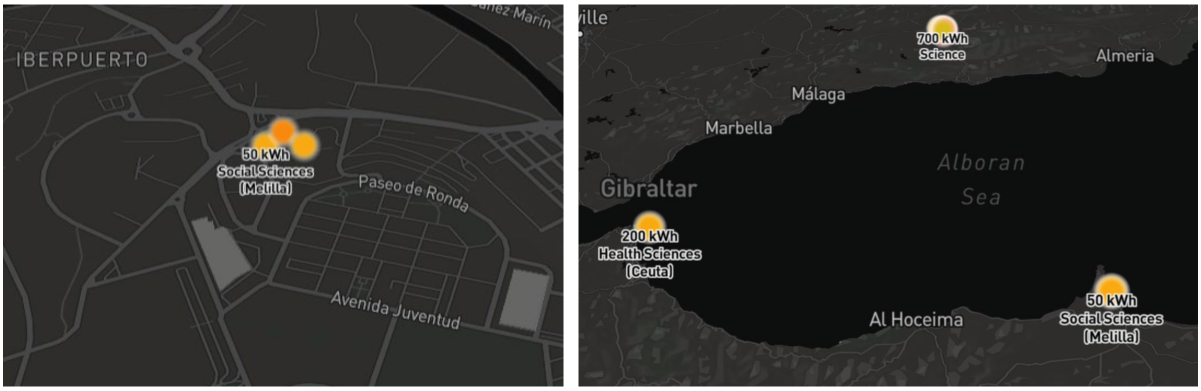


Fig. 4. Layer structure of the map visualization using Mapbox.

The last layer is designed to give an overview of the consumption in an aggregated way. Fig. 5a shows how it would be like if points were not clustered. In view of the low number of facilities in Ceuta and Melilla, buildings are relatively discernible in Fig. 5b and Fig. 5c. Nonetheless, given the amount of buildings located together, a fair number of them are overlapped in Fig. 5d. For all these reasons, the last layer has been configured to address these visualization problems. The next section will exemplify the final representation in detail.

(a) (b)



(c) (d)

Fig. 5. Overview of the buildings at different zoom levels. A) if no clustered is performed in the city of Granada with 24 buildings, in (b) the city of Ceuta with 2 facilities and in (c) Melilla with 3 Universities. In (d) the whole institution is depicted with all its buildings.

5. Results

Our dataset is made up of many buildings. As a consequence, a set of representative buildings were chosen to carry out all the experiments. Besides, to prevent an excessive amount of results, intermediate results will be skipped. Cross-validation were used to validate models, and each experiment was repeated 10 times and the average of those executions was taken in order to prevent biased results.

The first model, RT, was set with the next parameters: Standard CART as the algorithm used to select the best split predictor, the minimum number of branch node observations was 10 and the minimum leaf size was set to 36, the prune procedure enabled to estimate optimal sequence of pruned subtrees. Bagged and Boosted RT will be trained with 30 learners (minimum leaf size will be 8 observations). Secondly, SVMs were set using two different kernel functions: Linear and Gaussian. Hyper-parameters were optimized by following the Bull's method (Bull, 2011). Thirdly, the number of hidden layers of the ANN were considered as one layer with 10 neurons in both cases. ENN also used 10 feedback delays in the *context layer.* The learning algorithm will be Levenberg-Marquardt.

In order to evaluate the performance of the seven techniques, datasets are split into training set and testing set. Each model trained with 70% of the energy consumption data and the remaining 30% to measure the goodness of the fit.

According to all these directions, Table 1 was built; it shows the Root Mean Squared Error for all tested models: Regression Tree (RT), Bagged Regression Tree (BaRT), Boosted

Regression Tree (BoRT), Linear Support Vector Machine (L-SVM), Coarse Support Vector Machine (C-SVM), Multi-layer Perceptron (MLP) and Elman Neural Network (ENN). As mentioned above, eight buildings were chosen as representative examples of the whole dataset. The average RMSE of 10 executions is presented in Table 1. In addition to this, ten different models where designed so as to predict the following 10 hours. In this way, each column illustrates the RMSE obtained to predict the energy consumption within f hours, i.e., the first column shows the RMSE to predict the next hour, the second column shows the RMSE to predict the consumption within two hours, and so on.

At first glance, we can observe that all the models present a similar behavior in terms of performance except the four method, L-SVM. In the latter case, L-SVM seems to be the worst option to model this kind of data as it is, by far, the model with higher error in almost all cases. Only with the fifth building and predicting the $t + 1$ and $t + 2$ values it reaches a slightly lower error than its other version, C-SVM. On the basis of these results there appears to be a bad option to choose SVM as a forecasting method on this problem. Thus, the last place in the ranking is for L-SVM (7^{th}), followed by C-SVM (6^{th}). The fifth position in the ranking is among RG, BoRT and MLP. Interestingly, MLP better predicts in the near future, as it achieves better performance in case $f = 1$ and $f = 2$. However, when $f > 3$, MLP gets the 3^{rd} worst predictions in more than half of the cases (5^{th}). Besides, RT which is much simpler model than MLP accomplishes the 4^{th} position in the ranking in all instances except for the two first hours. The third position is reached by MLP only when predicting the next hour. In the rest of the cases this position is mainly held by BoRT (3^{rd}), followed by BaRT (2^{nd}) which attains better results on the order of a few hundredths. Note that MLP gets even the second best alternative for predicting in a few instances, such as B8 in all its prediction windows. Finally, when looking more in detail this table, one may see that the undisputed best model is ENN. Only 1 out of the 80 trials ENN gets the second best error as shown in building 6, $t + 1$. In this case, BaRT achieves an error of 0.0004 better than ENN. In spite of this, ENN predicts, by far, all the buildings and time windows much more accurate than the rest of models. In fact, ENN has an average RMSE of 25% better than the rest of the models, and in the best case it gets up to 45%.

Table 1. Root Mean Squared Error for Regression Tree, Bagged Tree, Boosted Tree, Linear Support Vector Machine, Coarse Support Vector Machine, Multi-layer Perceptron and Elman Neural Network; with a prediction window from 1 to 10 hours ahead.

Test	Forecasting $(t+f)$									
	1	2	3	4	5	6	7	8	9	10
Building 1										
RT	0.042423	0.054863	0.063985	0.070398	0.075528	0.078756	0.078588	0.079099	0.078691	0.079696
BaRT	0.037520	0.049720	0.057712	0.064939	0.070922	0.075486	0.075698	0.075596	0.075504	0.075297
BoRT	0.040604	0.052857	0.061667	0.068541	0.072794	0.075675	0.075981	0.076402	0.075726	0.075921
L-SVM	0.127427	0.178221	0.240546	0.217467	0.309545	0.443688	0.494352	0.465635	0.427360	0.528536
C-SVM	0.094915	0.105127	0.111173	0.118738	0.124602	0.124382	0.124944	0.125287	0.124864	0.124389
MLP	0.039017	0.058152	0.069397	0.078527	0.085269	0.088017	0.089041	0.089225	0.088262	0.088203
ENN	**0.034165**	**0.045475**	**0.051615**	**0.055554**	**0.056859**	**0.057094**	**0.059120**	**0.058332**	**0.059158**	**0.058138**
Building 2										
RT	0.042859	0.062125	0.076290	0.087660	0.098379	0.106283	0.110490	0.112531	0.114845	0.118492
BaRT	0.038226	0.056269	0.068648	0.079789	0.090390	0.097477	0.102186	0.105571	0.107050	0.109130
BoRT	0.044885	0.065514	0.078284	0.089700	0.099833	0.106933	0.110984	0.113611	0.116065	0.118391
L-SVM	0.102332	0.167063	0.212275	0.221770	0.266588	0.278696	0.310133	0.362435	0.328193	0.266761
C-SVM	0.054078	0.101398	0.125119	0.157197	0.178748	0.204213	0.218686	0.212897	0.222557	0.211742
MLP	0.035340	0.053955	0.067195	0.079754	0.089908	0.097028	0.102150	0.105771	0.106879	0.110309
ENN	**0.031797**	**0.044207**	**0.051942**	**0.057654**	**0.062532**	**0.063908**	**0.066410**	**0.066248**	**0.067870**	**0.067452**
Building 3										
RT	0.031547	0.049398	0.061997	0.073760	0.081238	0.089481	0.093897	0.097338	0.100127	0.102717
BaRT	0.026183	0.040338	0.051118	0.060697	0.067733	0.075033	0.079476	0.083319	0.085435	0.088874
BoRT	0.031496	0.046650	0.058416	0.066898	0.075262	0.082133	0.087292	0.090801	0.093283	0.096594
L-SVM	0.084617	0.128474	0.198676	0.259834	0.441548	0.393710	0.435010	0.510033	0.483196	0.481687
C-SVM	0.040866	0.075134	0.105658	0.148788	0.177875	0.232809	0.243370	0.253713	0.266257	0.278848
MLP	0.028075	0.045483	0.061151	0.074584	0.084634	0.094050	0.101026	0.105667	0.108644	0.111685
ENN	**0.023489**	**0.034806**	**0.042175**	**0.048729**	**0.051864**	**0.055996**	**0.057125**	**0.061281**	**0.059965**	**0.060444**
Building 4										
RT	0.032956	0.051007	0.063292	0.072987	0.080316	0.086200	0.091087	0.093816	0.095955	0.099187
BaRT	0.028625	0.041943	0.052018	0.061827	0.069075	0.075140	0.079385	0.081742	0.084069	0.087188
BoRT	0.035509	0.050631	0.061390	0.069367	0.076878	0.083751	0.087000	0.089930	0.092027	0.095168
L-SVM	0.103108	0.140093	0.252552	0.234440	0.327257	0.347307	0.364249	0.421843	0.351986	0.443102
C-SVM	0.053955	0.094886	0.134043	0.164638	0.173981	0.185513	0.198803	0.217501	0.218112	0.224203
MLP	0.032587	0.050652	0.065593	0.076973	0.086439	0.094264	0.098363	0.101248	0.102968	0.106977
ENN	**0.028453**	**0.041269**	**0.049153**	**0.053318**	**0.056964**	**0.059172**	**0.061813**	**0.062787**	**0.062372**	**0.063499**
Building 5										
RT	0.031764	0.042542	0.050314	0.055847	0.059425	0.062154	0.064093	0.065677	0.067083	0.068476
BaRT	0.028353	0.037966	0.044599	0.049644	0.053220	0.055963	0.057586	0.059398	0.061064	0.061911
BoRT	0.033847	0.042640	0.048179	0.052707	0.055439	0.057752	0.059338	0.060393	0.062129	0.062896
L-SVM	0.056076	0.072446	0.100790	0.110914	0.115419	0.131269	0.137492	0.122878	0.129351	0.146875
C-SVM	0.059594	0.079173	0.088712	0.099861	0.104707	0.106620	0.104460	0.107770	0.111743	0.098357
MLP	0.028927	0.041099	0.048657	0.055133	0.059068	0.062300	0.064611	0.066456	0.067604	0.068328
ENN	**0.026887**	**0.033301**	**0.035983**	**0.038896**	**0.040655**	**0.041526**	**0.041912**	**0.042513**	**0.043044**	**0.043061**
Building 6										
RT	0.027749	0.040425	0.048876	0.055236	0.060473	0.067483	0.074032	0.076152	0.078286	0.078564
BaRT	**0.022552**	0.032892	0.040235	0.045656	0.051210	0.056968	0.062204	0.065652	0.066764	0.068371
BoRT	0.026224	0.037051	0.043347	0.049336	0.055228	0.062078	0.067480	0.069971	0.071582	0.071924
L-SVM	0.125184	0.140588	0.149961	0.204499	0.289259	0.333380	0.358899	0.357593	0.366183	0.331850
C-SVM	0.072624	0.111474	0.133219	0.120124	0.120588	0.140640	0.142243	0.146175	0.145976	0.147139
MLP	0.031053	0.042480	0.050606	0.057242	0.063189	0.071036	0.078087	0.081710	0.083372	0.085115
ENN	0.022911	**0.029707**	**0.033855**	**0.034691**	**0.037531**	**0.038591**	**0.039204**	**0.039534**	**0.039872**	**0.040486**
Building 7										
RT	0.030488	0.049363	0.064206	0.079134	0.090885	0.099889	0.106828	0.110788	0.114986	0.118440
BaRT	0.025758	0.039776	0.052577	0.064917	0.074822	0.083428	0.089966	0.094206	0.098149	0.101023
BoRT	0.033699	0.050512	0.062146	0.074838	0.084957	0.092793	0.098470	0.102427	0.106219	0.109829
L-SVM	0.111623	0.154890	0.261622	0.327596	0.482860	0.416295	0.359808	0.420238	0.419129	0.484455
C-SVM	0.050886	0.096670	0.131943	0.192059	0.192439	0.211648	0.232536	0.270985	0.268914	0.281900
MLP	0.027762	0.047868	0.066254	0.081760	0.095626	0.105703	0.113146	0.116764	0.120329	0.125132
ENN	**0.022944**	**0.033725**	**0.042692**	**0.047951**	**0.052137**	**0.055154**	**0.057240**	**0.056857**	**0.057524**	**0.058780**
Building 8										
RT	0.045387	0.066067	0.082481	0.096021	0.108163	0.117628	0.121427	0.122936	0.124140	0.126143
BaRT	0.039959	0.059244	0.074879	0.087520	0.099768	0.109324	0.113431	0.114853	0.116204	0.118459
BoRT	0.049992	0.070498	0.085710	0.099173	0.111072	0.118904	0.121412	0.123020	0.124214	0.125712
L-SVM	0.124275	0.210009	0.202029	0.234078	0.415058	0.387901	0.443007	0.386421	0.403097	0.479505
C-SVM	0.067852	0.107659	0.138375	0.170347	0.192552	0.217577	0.225939	0.225918	0.225650	0.221182
MLP	0.038648	0.057840	0.072436	0.084582	0.096046	0.104135	0.107953	0.109543	0.110910	0.112472
ENN	**0.031425**	**0.043970**	**0.050122**	**0.056304**	**0.060288**	**0.062305**	**0.063147**	**0.062334**	**0.062709**	**0.062508**

In order to confirm the results that have just been discussed, Table 2 gathers the p-values obtained from the statistical t-test. Since the ENN was the most accurate model, it was compared with the rest of the models. Thus, the test returns a decision for the null hypothesis that the data in ENN and in the model to be compared comes from distributions with equal means. Furthermore, to avoid an unnecessary big table, we set the future value to the next value, i.e., $t + 1$. Thus, one can see that the table supports the previous assertions; there are significant differences in the ENN predictor in nearly all the cases but in building 4 and 6 in whose cases they do not exist at 95% confidence. Besides, one can observe this fact in the previous table and how both errors were very similar.

Table 2. Result of the statistical t-test comparing two samples, the ENN versus the rest of the models.

Building	Model					
	RT	BaT	BoT	L-SVM	C-SVM	MLP
B1	2.547E-13	1.947E-06	1.807E-11	3.573E-32	3.459E-29	6.242E-08
B2	2.483E-15	1.279E-11	4.175E-18	1.585E-29	1.577E-20	3.339E-08
B3	5.626E-11	1.168E-04	3.067E-11	7.934E-27	3.551E-18	1.232E-07
B4	1.830E-08	**6.260E-01**	8.586E-11	4.599E-29	2.290E-20	4.633E-08
B5	1.526E-08	4.813E-03	3.343E-11	5.060E-22	5.676E-23	1.156E-03
B6	1.656E-08	**3.419E-01**	4.563E-06	3.412E-31	3.718E-25	1.150E-11
B7	3.266E-11	2.255E-06	1.426E-15	1.408E-31	2.558E-21	1.946E-08
B8	3.680E-15	2.135E-11	1.154E-16	3.673E-30	1.105E-22	4.291E-10

The best results obtained are summarized in Table 3. As has just been discussed, ENN is the best approach to model the energy consumption in this case study. When looking in detail this table, one may observe that errors increase as the forecasting time step dilates. To better observe this particular behaviour, Fig. 6 shows the fact of the RMSE increase while the values to be predicted are further in time. In the case of ENN, this increase is slightly softer if compared to all the other models.

Table 3. Root Mean Squared Error of the best prediction model with different forecasting time steps.

Building	Forecasting $(t + f)$									
	1	2	3	4	5	6	7	8	9	10
B1	0.034165	0.045475	0.051615	0.055554	0.056859	0.057094	0.059120	0.058332	0.059158	0.058138
B2	0.031797	0.044207	0.051942	0.057654	0.062532	0.063908	0.066410	0.066248	0.067870	0.067452
B3	0.023489	0.034806	0.042175	0.048729	0.051864	0.055996	0.057125	0.061281	0.059965	0.060444
B4	0.028453	0.041269	0.049153	0.053318	0.056964	0.059172	0.061813	0.062787	0.062372	0.063499
B5	0.026887	0.033301	0.035983	0.038896	0.040655	0.041526	0.041912	0.042513	0.043044	0.043061
B6	0.022552	0.029707	0.033855	0.034691	0.037531	0.038591	0.039204	0.039534	0.039872	0.040486
B7	0.022944	0.033725	0.042692	0.047951	0.052137	0.055154	0.057240	0.056857	0.057524	0.058780
B8	0.031425	0.043970	0.050122	0.056304	0.060288	0.062305	0.063147	0.062334	0.062709	0.062508

To further support this last observation, we studied whether any significant difference was manifested in the different time steps analysed and those results are shown in Table 4. In this table, each row stands for the time step to be analyzed and the column amounts to the time step to be compared. Obviously, the p-value for the values of the diagonal are 1 as they compare the same time step. This table also illustrates how the smaller is the time step, the greater is the difference. However, there is no significant differences among time steps larger than 5. This leads us to conclude that the model stabilizes with consistent predictions after that point.

Table 4. Results of the statistical t-test for the forecasting time step comparing all versus all.

	Forecasting $(t+f)$									
	1	2	3	4	5	6	7	8	9	10
1	**1.000E+00**	6.695E-13	7.971E-16	9.956E-17	1.943E-16	4.541E-17	1.293E-18	7.946E-18	4.555E-17	8.419E-19
2	6.695E-13	**1.000E+00**	3.337E-09	6.287E-12	5.800E-12	9.673E-13	5.218E-15	6.645E-14	3.477E-13	4.448E-15
3	7.971E-16	3.337E-09	**1.000E+00**	1.673E-05	1.707E-06	3.202E-07	3.839E-10	7.527E-09	1.161E-08	1.129E-09
4	9.956E-17	6.287E-12	1.673E-05	**1.000E+00**	**1.234E-01**	**5.623E-02**	5.353E-05	1.201E-03	3.455E-04	7.669E-04
5	1.943E-16	5.800E-12	1.707E-06	**1.234E-01**	**1.000E+00**	**7.786E-01**	**7.738E-03**	**8.046E-02**	**1.797E-02**	**9.233E-02**
6	4.541E-17	9.673E-13	3.202E-07	**5.623E-02**	**7.786E-01**	**1.000E+00**	**9.491E-03**	**1.125E-01**	**2.369E-02**	**1.311E-01**
7	1.293E-18	5.218E-15	3.839E-10	5.353E-05	**7.738E-03**	**9.491E-03**	**1.000E+00**	**2.510E-01**	**9.612E-01**	**1.033E-01**
8	7.946E-18	6.645E-14	7.527E-09	1.201E-03	**8.046E-02**	**1.125E-01**	**2.510E-01**	**1.000E+00**	**3.198E-01**	**7.604E-01**
9	4.555E-17	3.477E-13	1.161E-08	3.455E-04	**1.797E-02**	**2.369E-02**	**9.612E-01**	**3.198E-01**	**1.000E+00**	**1.813E-01**
10	8.419E-19	4.448E-15	1.129E-09	7.669E-04	**9.233E-02**	**1.311E-01**	**1.033E-01**	**7.604E-01**	**1.813E-01**	**1.000E+00**

On the other hand, if we continue observing Fig. 6, we can distinguish that there are buildings whose error is moderately lower than the rest of the buildings. This fact is on account of the particularities of the building as there are buildings whose behaviour are less changing and fluctuating than others, and therefore those have a more foreseeable energy consumption. In any case, as shown in Fig. 6 and Table 3, ENN's errors are relatively close to one another. Besides, when analysing in detail these results and taking $t+1$ as a reference point, the next predictions are worsening by 38% for $t+2$, on average. In the next case, it reaches an increase of 62%, 78% and 90% for $t=\{3,4,5\}$ respectively. After that, the error passes with a difference of 100% and indicates a smooth growth from that point.

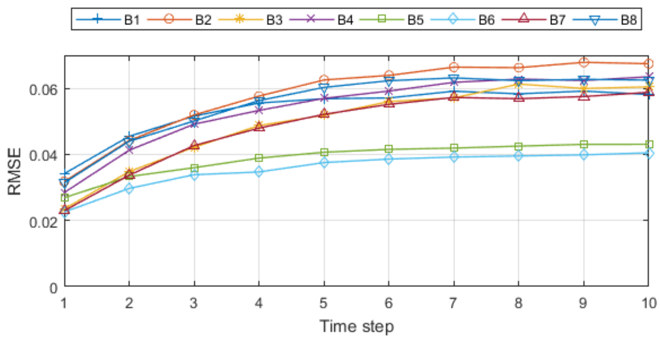


Fig. 6. Evolution of the RMSE as the forecasting time step increases.

Once forecasting models have been properly analyzed and the best option is selected, the need to go one step further is essential. As a result, the rest of this section will describe how VIOMEN was deployed at UGR to visualize the energy consumption at its campuses along with the related predictions provided by the forecasting models.

An initial appearance of the application is depicted in Fig. 7. As seen in the figure, two-thirds of the screen are dedicated to the map visualization where buildings are displayed, which is where most of the relevant knowledge will be presented. This area also contains two buttons (located in the upper left corner of the view) used to modify clustering and colour properties, and a textbox which allows us to look for places and streets on the map by typing global search terms. The remaining one-third of the application (right part) lists all the buildings, serving as a "calling card" of each building.

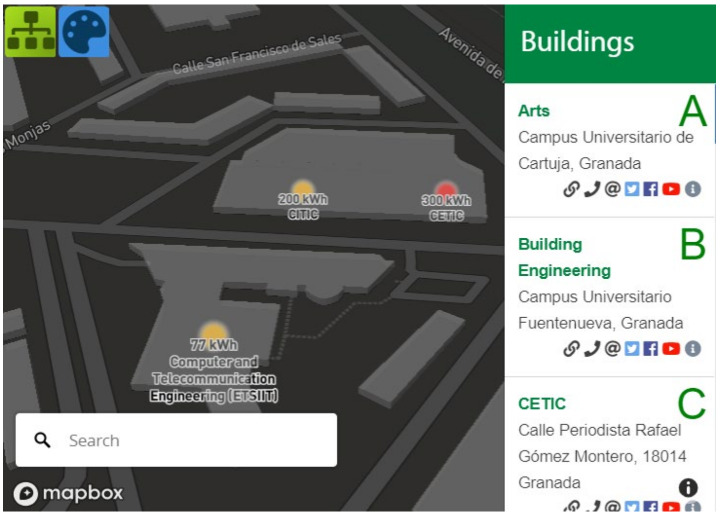


Fig. 7. Example of a general view of the proposed system.

Due to the relatively high number of co-located elements when zooming out, it is necessary to perform a building grouping. Thus, VIMOEN reduces the number of drawn points on the grid when points collide. It is possible to achieve various forms of aggregation, and we consider four of them (as illustrated on Fig. 8a).

A cluster's energy consumption may be represented by the maximum, the minimum, the average or the total amount of the consumption in that group. The purpose of these adjustments is to adapt the system to user needs. For instance, the maximum value allows easy detection of a potential building problem in the consumption; alternatively, the measure of the average gives an idea of the consumption for each of the areas concerned; similarly, the other two options furnish valuable information depending on the context from which the subject is approached.

In keeping with this philosophy, the colour of the clusters is also considered flexible to adjustments. To this end, the colour of the group admits three strategies as Fig. 8b illustrates. The mixture option performs the averaging of the red, green and blue components and the alpha channel. The high and low options depicts the group according to the biggest and smallest consumption respectively. In doing so, the user can see a quick description of the whole campus by checking the colours of the buildings. For instance, by choosing «Max» as the aggregate property and «Mixture» as the colour of the cluster (see Fig. 8) the user is capable of identifying groups in which there are buildings that have the highest consumption and therefore possible problems in them. On the other hand, with this option selected, if there are no reddish buildings, then we can assure that there are no issues in any building.

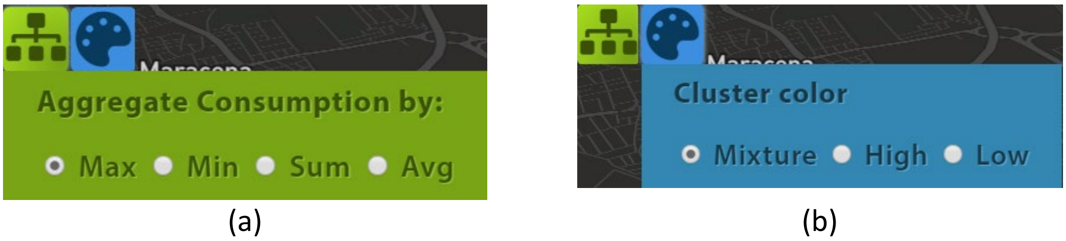


Fig. 8. Upper left buttons: available (a) clustering properties and (b) colour behaviour of the group.

We have so far described VIMOEN's colour properties and the diverse ways of visually summarizing the data. As the consumption profiles of each building are dissimilar, each point should be coloured differently according to the range of its historical energy use. In other words, there is a colour palette where green, orange and red represent that the building's consumption is close to its recorded minimum, medium and maximum energy values respectively. Using these three colours, a set range of 10 colours have been created to distinguish the consumption considering the energy use of each building. Thereby, two buildings with the same consumption at a certain time could have different colour, because their historical consumptions are different. There is a similar situation in Fig. 7 where two buildings with different consumption are coloured with the same colour due to the different range of consumption.

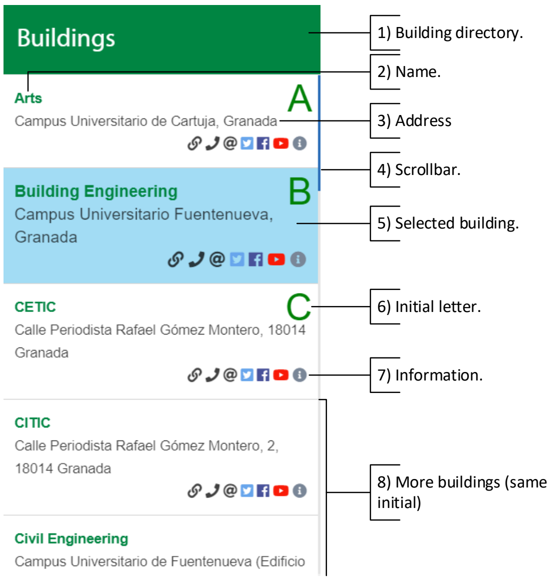


Fig. 9. Partial view of the list of UGR's buildings in VIMOEN.

Fig. 9 shows the area listing the existing buildings. The building directory is sorted alphabetically by name for convenience. Buildings' names are distinguished by its initial letter. Each description card summarizes information about one building, e.g., the name of the building, its specific address (street, number, city), the URL of the building website, the contact phone number, an e-mail address, its social networks (Twitter, Facebook and YouTube) and other information that describes the building. The design of the menu is a simple interface for showing and locating the buildings in the map quickly and straightforwardly. Consequently, as we were mentioning above, if there were any problem with a specific building, the person responsible for monitoring the buildings might send an e-mail or even call to that building to report about a potential problem.

Fig. 10. Example of map pin when building is select in the list.

Thus, a white map pin appears when the mouse pointer is placed into the building card from the list and its cell is displayed in light blue. This is reflected in Fig. 10. What's more, the application incorporates the functionality of adjusting the current location to the new selected building. This is carried out when click on one of the rows of the list, providing for easy location of the buildings.

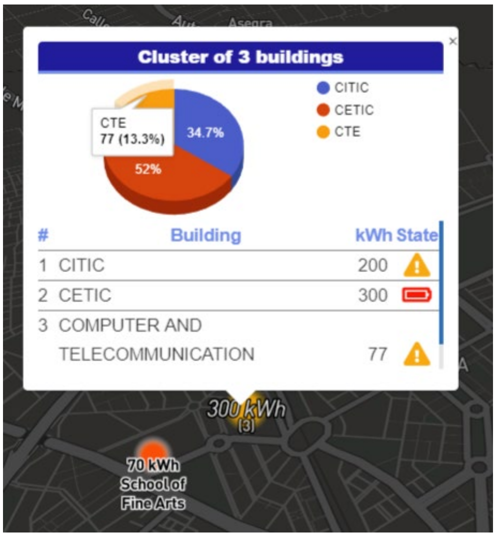


Fig. 11. Example of a popup for a cluster.

Once a building (or cluster of them) is selected, a popup appears with useful energy consumption information. Fig. 11 details the information for a building group of three edifices. The proportional amount of energy from each building is represented with an SVG-rendered pie chart. When pieces of the chart are selected, this graph also provides detailed individual figures. In order to optimize the space of the chart, the name of the buildings is labelled by taking its initials, e.g., the *CTE* building is the Computer and Telecommunications Engineering School. Tooltips are also displayed on hovering. By looking at this information, one can check whether the ratio of energy usage makes sense or not because bigger buildings, more often than not, will have larger consumption. Fig. 9 illustrates three buildings, the first and second buildings are a research centre and a building to host small and medium-size enterprises of the UGR. The third building is a teaching centre. Let's assume that the first two buildings are smaller than the third one. In this case, the third building presents the lower consumption, and this fact sets off the alarms because something wrong is happening here. At this point, two things could be happening: 1) CITIC and CETIC are consuming a disproportionate amount of energy or 2) CTE presents an unexpected low consumption. However, this event could be explained by the current period, in other words, this could be caused by an exam period or holidays, otherwise, the manager might want to contact the building operator for further details.

In addition to this, a description of the consumption appears below the graph. All buildings are listed using its complete name and its current energy consumption. The last column of the scrollable list gives visual information about the building's behaviour. As illustrated before in Fig. 7, two of the three buildings are consuming above their average, and the other one is consuming close to its historical maximum. Fig. 12a lists all available icons, together with their meaning, for visuallysummarizing the energy behaviour of the building.

	Consumption below its average.
	It is consuming beyond the average.
	Consumption near the max recorded.
	Energy consumption unavailable.

(a)

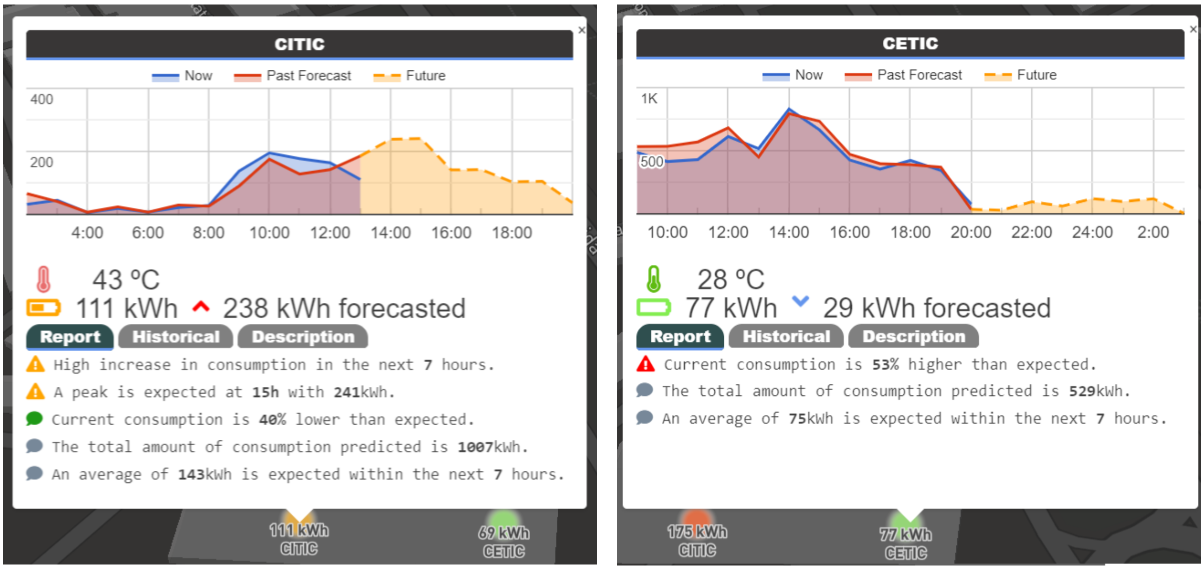
	The consumption is predicted to increase.
	The consumption is predicted to decrease.
	The consumption maintains its value.

(b)

Fig. 12. Set of icons for representing the (a) current status and (b) the future consumption of the buildings.

Apart from that, the individual popup describes important information about a building (Fig. 13). The specific popup of one building displays the current energy consumption on the central left-hand together with a battery icon which indicates the state of the building according to the range of energy use. The colour and the filling of the battery change dynamically according to current consumption. The predicted consumption is located to the right of the current consumption. This information is also presented with an animation (Fig. 12b) indicating if the prediction goes up, down or maintains its value. The specific forecasting value is shown next to that animated icon. Not only the current and the forecasting consumption is reflected in the panel, but also a historical evolution of the building (blue line), the past forecasted values (red) and the future consumption (orange). In this way, the user may observe the evolution of the actual consumption in the recent past, the past forecasting values help the user to discern if the consumption at that period has been expected or, on the contrary, it differs from the predictions. In addition to this, the user can anticipate future behaviours in buildings by using all the foreseen information from forecasting models. The orange time-series describes an easy way to see at a glance what is going to take place. Furthermore, at the bottom of the panel, there are three tabs containing relevant information about the building. For instance, Fig. 13a depicts a warning as the consumption will grow considerably, as well as an expected peak within 2 hours. In this case, the current consumption was lower than predicted, so the system reports it as well, along with the total consumption and an average of the consumption per hour, although each specific prediction may be found in the main chart. On the contrary, Fig. 13b depicts the energy consumption of a building whose actual consumption is close to its minimum (green empty battery), and it is expected to decrease its consumption in the next hour (blue arrow). In this building, an alert is shown as the current consumption is significantly higher than expected. Nevertheless, there are no warnings about peaks nor high increases, as in Fig. 13a.

All this information can be used to prevent the waste of energy. For example, Fig. 13a depicts a high increase within the next 7 hours. These predictions are not only informative but they are describing a behaviour that seems to be common in that building. So, we should analyse why is this, and whether this consumption is appropriate for that building. Besides, this chart illustrates a two-fold increase in the next hour, at 14:00, and a peak in the following hour, at 15:00. But this is not only a simple increase: it also exposes the consumption in the evening is beyond the morning's. This fact could be trivial but in Spain the working time is usually in the morning/afternoon, and the lunch time is between 14:00 and 15:00. With this in mind, we could formulate some hypotheses like that buildings could be consuming inefficiently at least that day. The user should gather all the information available and report about this to enhance the energy efficiency of that building. An assumption of what is going on there could be that the people who were in the morning in that building could have left their air conditioner switched on and those to come in the evening will also turn their heating on, increasing the global consumption of the building.



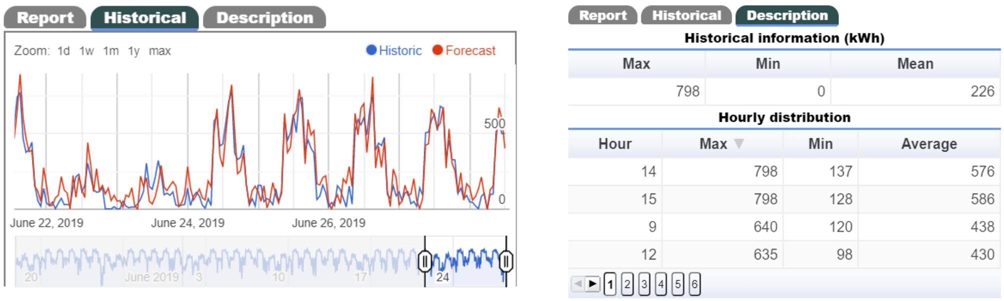
(a) (b)

Fig. 13. Example of the individual popup for two buildings with different information, alerts and energy consumption range.

Lastly, some buildings also collect the temperature conditions, that information is depicted with a thermometer which is filled in accordance with the registered maximum and minimum temperature of the building, in the same way as the battery icon. Both icons flash in red when they are close to the maximum value.

A final feature of VIMOEN is illustrated in Fig. 14. It is essential to offer the user the maximum information possible, as a consequence, the historical consumption could be examined. This feature is made up of two parts, the whole consumption stored on the bottom, and the selected period which is depicted on the top. And to fully exploit the data, a historical report about the consumption is rendered in the third tab (Fig. 14b). This tab gathers the maximum, minimum and the mean recorded up to now, along with the detailed distribution of the consumption in all hours, with the possibility to sort them out to make the most of the depicted information, e.g., at 14:00 is the period with the higher maximum.

Up to now, we have described some examples and situations that VIMOEN can be used to analyse, detect and enhance energy efficiency in buildings in real-time. Having said this, the historical data is essential also to learn about the consumption and the usual behaviour of the buildings as well as the fact that this register can be combined with the status that VIMOEN is providing about the buildings at a particular time. As just commented, we can draw conclusions about the period that takes more energy and the period and the lesser one. Besides, we can identify whether there is any problem with the predictive models by checking the forecasting series (red line) against the registered consumption (blue line). If big differences between those two series are found, then the forecasting models can be predicting improperly,. In addition to this, the historical data can be used to analyse patterns in the consumption (Fig. 14a demonstrates this fact). By focusing on the last week, we can see that the previous four and a half of the days present similar behaviour, also it seems to be that the first part of the graph would correspond to rest of the day in which we currently are, so we can get an idea of what will be the consumption soon. And finally, we can distinguish that there are two days which have lower consumption and have different pattern.



(a) (b)

Fig. 14. Example of (a) the historical consumption (focusing on the last week) and (b) description of the hourly distribution energy consumption.

The combination of VIMOEN with the experience, skills and dedication of the users produces a very powerful tool which is useful and helpful for monitoring and management of the energy consumption in distributed facilities.

6. Conclusion

This work describes a complete methodology to solve the problem of energy consumption forecasting along with the many-times forgotten application in real life. In our study, we apply data pre-processing to transform data from sensors into a more usable way which includes noisy treatment and data imputation. Our experimentation compares several forecasting models in order to obtain the best approach. In this study Elman Neural Network has proven to be the most accurate model to predict energy consumption. In addition to this, different time steps were tested to enrich and improve the functionality of the system, as well as to confirm the validity of the model to make medium-term predictions. Thus, the best prediction model is nominated to make predictions in VIMOEN, a software tool for energy visualization in institutions and organizations whose facilities are distributed throughout several geographic zones.

VIMOEN is designed to easily incorporate new facilities and updated knowledge about new structures. As such, it is a novel prototype for visualization and monitoring the energy consumption which comprises the visualization of historical data with a view to detect and control energy problems for the future, with the intention of contributing and assisting in decision-making processes, and therefore, enabling everyday saving in energy consumption (reported in real time as shown in Fig. 14a). What’s more, an advantage of our design is its simplicity. However, this strength limits the wealth of knowledge which could be displayed, since it is harder to understand dashboards that are overstuffed with information. Our choice is made to ensure an intuitive, easy-to-use interface and suitable representation of the knowledge to assist user interaction with the system.

VIMOEN provides a solution for examining and displaying the energy consumption of a set of geographically-distributed buildings. This task, which is costly, complex, time-consuming and usually carried out manually using spreadsheets, changes into an easy visual exercise thanks to VIMOEN. As it enables the visualization of all distributed buildings of an organization and incorporates their past, current and future consumptions.

For its implementation, we used Mapbox, which allows the creation of three-dimensional (3D) effects with the fill extrusion layer. Even though building models with a high level of details can be a costly and time-consuming process, it could be a very useful approach for future work. Thus, future research will be oriented at developing three-dimensional (3D) models of the buildings with a low level of detail. And indeed there are some recent methodologies for the creation of mock-up 3D models of buildings using an image-based automated process (Oskouie, Becerik-Gerber & Soibelman, 2017). This would contribute to creating a more accurate view of the whole campus. Along these lines, the operating rate of a building could be considered for visualization in those 3D models and thus serve to facilitate cost control, distinguishing between busy and inactive spaces. And besides, it would improve the energy prediction because of some researches point out that the occupancy is the most important factor affecting energy waste (Deb, Eang, Yang & Santamouris, 2016).

7. Abbreviations

ANN	Artificial Neural Network.
BaRT	Bagged Regression Tree.
BoRT	Boosted Regression Tree.
C-SVM	Coarse Support Vector Machine.
ENN	Elman Neural Network
L-SVM	Linear Support Vector Machine.
MLP	Multi-Layer Perceptron.
RMSE	Root Mean Squared Error.
RT	Regression Tree.
SVM	Support Vector Machine.

8. Acknowledgments

This work has been developed with the support of the Department of Computer Science and Artificial Intelligence of the University of Granada, Approximate Reasoning and Artificial Intelligence Group (TIC111), and the project TIN201564776-C3-1-R. MMS was supported by European Union's H2020 Athenea3i programme (ga. No. 754446).

9. Bibliography

Ahn, J. & Cho, S. (2017). Dead-band vs. machine-learning control systems: Analysis of control benefits and energy efficiency. *Journal of Building Engineering, 12*, 17-25. http://dx.doi.org/10.1016/j.jobe.2017.04.014

Andiojaya, A. & Demirhan, H. (2019). A bagging algorithm for the imputation of missing values in time series. *Expert Systems with Applications, 129*, 10-26. http://dx.doi.org/10.1016/j.eswa.2019.03.044

Aznavi, S., Fajri, P., Sabzehgar, R. & Asrari, A. (2020). Optimal management of residential energy storage systems in presence of intermittencies. *Journal of Building Engineering, 29*, 101149. http://dx.doi.org/10.1016/j.jobe.2019.101149

Bilous, I., Deshko, V. & Sukhodub, I. (2018). Parametric analysis of external and internal factors influence on building energy performance using non-linear multivariate regression models. *Journal of Building Engineering, 20*, 327-336. http://dx.doi.org/10.1016/j.jobe.2018.07.021

Bull, A. D. (2011). Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research, 12*, 2879-2904. Retrieved from http://arxiv.org/abs/1101.3501v3

Chen, J., Shi, Q., Shen, L., Huang, Y. & Wu, Y. (2019). What makes the difference in construction carbon emissions between China and USA? *Sustainable Cities and Society, 44*, 604-613. http://dx.doi.org/10.1016/j.scs.2018.10.017

Davlea, L. & Teodorescu, B. (2016). A neuro-fuzzy algorithm for middle-term load forecasting. In *2016 International Conference and Exposition on Electrical and Power Engineering (EPE)* (pp. 005-009). http://dx.doi.org/10.1109/ICEPE.2016.7781292

Deb, C., Eang, L. S., Yang, J. & Santamouris, M. (2016). Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks. *Energy and Buildings, 121*, 284-297. http://dx.doi.org/10.1016/j.enbuild.2015.12.050

Deb, C., Zhang, F., Yang, J., Lee, S. E. & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews, 74*, 902-924. http://dx.doi.org/10.1016/j.rser.2017.02.085

Duque-Pintor, F., Fernández-Gómez, M., Troncoso, A. & Martínez-Álvarez, F. (2016). A New Methodology Based on Imbalanced Classification for Predicting Outliers in Electricity Demand Time Series. *Energies, 9*, 752-752. http://dx.doi.org/10.3390/en9090752

Elbeltagi, E., Wefki, H., Abdrabou, S., Dawood, M. & Ramzy, A. (2017). Visualized strategy for predicting buildings energy consumption during early design stage using parametric analysis. *Journal of Building Engineering, 13*, 127-136. http://dx.doi.org/10.1016/j.jobe.2017.07.012

Fu, T.-c. (2011). A review on time series data mining. *Engineering Applications of Artificial Intelligence, 24*, 164-181. http://dx.doi.org/10.1016/j.engappai.2010.09.007

Fumo, N. & Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. *Renewable and Sustainable Energy Reviews, 47*, 332-343. http://dx.doi.org/10.1016/j.rser.2015.03.035

García-Hinde, O., Terrén-Serrano, G., Hombrados-Herrera, M. Á., Gómez-Verdejo, V., Jiménez-Fernández, S., Casanova-Mateo, C., Sanz-Justo, J., Martínez-Ramón, M. & Salcedo-Sanz, S. (2018). Evaluation of dimensionality reduction methods applied to numerical weather models for solar radiation forecasting.

Engineering Applications of Artificial Intelligence, 69, 157-167. http://dx.doi.org/10.1016/j.engappai.2017.12.003

Gelazanskas, L. & Gamage, K. A. A. (2014). Demand side management in smart grid: A review and proposals for future direction. *Sustainable Cities and Society, 11*, 22-30. http://dx.doi.org/10.1016/j.scs.2013.11.001

Gómez-Romero, J., Fernández-Basso, C. J., Cambronero, M. V., Molina-Solana, M., Campaña, J. R., Ruiz, M. D. & Martin-Bautista, M. J. (2019). A Probabilistic Algorithm for Predictive Control With Full-Complexity Models in Non-Residential Buildings. *IEEE Access, 7*, 38748-38765. http://dx.doi.org/10.1109/ACCESS.2019.2906311

Granderson, J. & Lin, G. (2016). Building energy information systems: synthesis of costs, savings, and best-practice uses. *Energy Efficiency, 9*, 1369-1384. http://dx.doi.org/10.1007/s12053-016-9428-9

Huacón, C. F. & Pelegrin, L. (2017). SURV: A system for massive urban data visualization. In *Undergraduate Research Technology Conference (URTC), 2017 IEEE MIT* (pp. 1-4): IEEE. http://dx.doi.org/10.1109/URTC.2017.8284174

International Energy Agency. (2013). Transition to Sustainable Building. Strategies and Opportunities to 2050. In. Retrieved from http://www.iea.org/publications/freepublications/publication/transition-to-sustainablebuildings.html

Iruela, J. R. S., Ruiz, L. G. B., Pegalajar, M. C. & Capel, M. I. (2020). A parallel solution with GPU technology to predict energy consumption in spatially distributed buildings using evolutionary optimization and artificial neural networks. *Energy Conversion and Management, 207*, 112535. http://dx.doi.org/10.1016/j.enconman.2020.112535

Jung, H. C., Kim, J. S. & Heo, H. (2015). Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach. *Energy and Buildings, 90*, 76-84. http://dx.doi.org/10.1016/j.enbuild.2014.12.029

Kayacan, E., Ulutas, B. & Kaynak, O. (2010). Grey system theory-based models in time series prediction. *Expert Systems with Applications, 37*, 1784-1789. http://dx.doi.org/10.1016/j.eswa.2009.07.064

Leon-Alcaide, P., Rodriguez-Benitez, L., Castillo-Herrera, E., Moreno-Garcia, J. & Jimenez-Linares, L. (2020). An evolutionary approach for efficient prototyping of large time series datasets. *Information Sciences, 511*, 74-93. http://dx.doi.org/10.1016/j.ins.2019.09.044

Lopes, M. A. R., Antunes, C. H. & Martins, N. (2012). Energy behaviours as promoters of energy efficiency: A 21st century review. *Renewable and Sustainable Energy Reviews, 16*, 4095-4104. http://dx.doi.org/https://doi.org/10.1016/j.rser.2012.03.034

Luzi, M., Vaccarini, M. & Lemma, M. (2019). A tuning methodology of Model Predictive Control design for energy efficient building thermal control. *Journal of Building Engineering, 21*, 28-36. http://dx.doi.org/10.1016/j.jobe.2018.09.022

MapBox & LLC. (2018). Mapbox GL JS API. *URL http://www.mapbox.com/mapbox-gl-js/api/*.

Martínez-Municio, S., Rodríguez-Benítez, L., Castillo-Herrera, E., Giralt-Muiña, J. & Jiménez-Linares, L. (2018). Linguistic Modeling and Synthesis of Heterogeneous Energy Consumption Time Series Sets. *International Journal of Computational Intelligence Systems, 12*, 259-272. http://dx.doi.org/10.2991/ijcis.2018.125905639

Molina-Solana, M., Ros, M., Ruiz, M. D., Gómez-Romero, J. & Martin-Bautista, M. J. (2017). Data science for building energy management: A review. *Renewable and Sustainable Energy Reviews, 70*, 598-609. http://dx.doi.org/10.1016/j.rser.2016.11.132

Moreno-Garcia, J., Rodriguez-Benitez, L., Jimenez-Linares, L. & Triviño, G. (2019). A Linguistic Extension of Petri Nets for the Description of Systems: An Application to Time Series. *IEEE Transactions on Fuzzy Systems, 27*, 1818-1832. http://dx.doi.org/10.1109/TFUZZ.2019.2892340

Murshed, S., Al-Hyari, A., Wendel, J. & Ansart, L. (2018). Design and Implementation of a 4D Web Application for Analytical Visualization of Smart City Applications. *ISPRS International Journal of Geo-Information, 7*, 276. Retrieved from http://www.mdpi.com/2220-9964/7/7/276

Nagy, G. I., Barta, G., Kazi, S., Borbély, G. & Simon, G. (2016). GEFCom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. *International Journal of Forecasting, 32*, 1087-1093. http://dx.doi.org/10.1016/j.ijforecast.2015.11.013
Najjar, M., Figueiredo, K., Palumbo, M. & Haddad, A. (2017). Integration of BIM and LCA: Evaluating the environmental impacts of building materials at an early stage of designing a typical office building. *Journal of Building Engineering, 14*, 115-126. http://dx.doi.org/10.1016/j.jobe.2017.10.005
Natconsumers. (2015). Natural language energy for promoting sustainable consumer behaviour. *H2020-EU.3.3.6.*, Retrieved from http://cordis.europa.eu/project/id/657672
Oh, T. K., Lee, D., Park, M., Cha, G. & Park, S. (2018). Three-Dimensional Visualization Solution to Building-Energy Diagnosis for Energy Feedback. *Energies, 11*, 1736. Retrieved from http://www.mdpi.com/1996-1073/11/7/1736
Oskouie, P., Becerik-Gerber, B. & Soibelman, L. (2017). Automated Recognition of Building Façades for Creation of As-Is Mock-Up 3D Models. *Journal of Computing in Civil Engineering, 31*, 04017059. http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000711
Pahl, S., Goodhew, J., Boomsma, C. & Sheppard, S. R. (2016). The role of energy visualization in addressing energy use: Insights from the eViz project. *Frontiers in psychology, 7*, 92. Retrieved from http://www.frontiersin.org/articles/10.3389/fpsyg.2016.00092/full
Pérez-Lombard, L., Ortiz, J. & Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings, 40*, 394-398. http://dx.doi.org/10.1016/j.enbuild.2007.03.007
Prasad, A. M., Iverson, L. R. & Liaw, A. (2006). Newer Classification and Regression Tree Techniques: Bagging and Random Forests for Ecological Prediction. *Ecosystems, 9*, 181-199. http://dx.doi.org/10.1007/s10021-005-0054-1
Ribeiro, G. T., Mariani, V. C. & Coelho, L. d. S. (2019). Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Engineering Applications of Artificial Intelligence, 82*, 272-281. http://dx.doi.org/10.1016/j.engappai.2019.03.012
Rodrigues, N., Netzel, R., Ullah, K. R., Burch, M., Schultz, A., Burger, B. & Weiskopf, D. (2017). Visualization of time series data with spatial context: communicating the energy production of power plants. In *Proceedings of the 10th International Symposium on Visual Information Communication and Interaction* (pp. 37-44): ACM. http://dx.doi.org/10.1145/3105971.3105982
Ruiz, L. G. B. (2020). Gestión de la Eficiencia Energética Mediante Técnicas de Minería de Datos. Retrieved from http://hdl.handle.net/10481/58703
Ruiz, L. G. B., Capel, M. I. & Pegalajar, M. C. (2019). Parallel memetic algorithm for training recurrent neural networks for the energy efficiency problem. *Applied Soft Computing, 76*, 356-368. http://dx.doi.org/10.1016/j.asoc.2018.12.028
Ruiz, L. G. B., Cuellar, M. P., Delgado, M. & Pegalajar, M. C. (2016). An Application of Non-Linear Autoregressive Neural Networks to Predict Energy Consumption in Public Buildings. *Energies, 9*, 21. http://dx.doi.org/10.3390/en9090684
Ruiz, L. G. B., Rueda, R., Cuéllar, M. P. & Pegalajar, M. C. (2018). Energy consumption forecasting based on Elman neural networks with evolutive optimization. *Expert Systems with Applications, 92*, 380-389. http://dx.doi.org/10.1016/j.eswa.2017.09.059
Runge, J., Zmeureanu, R. & Le Cam, M. (2020). Hybrid short-term forecasting of the electric demand of supply fans using machine learning. *Journal of Building Engineering, 29*, 101144. http://dx.doi.org/10.1016/j.jobe.2019.101144
Santamouris, M. & Kolokotsa, D. (2015). On the impact of urban overheating and extreme climatic conditions on housing, energy, comfort and environmental quality of vulnerable population in Europe. *Energy and Buildings, 98*, 125-133. http://dx.doi.org/10.1016/j.enbuild.2014.08.050
Sarma, P., Singh, H. K. & Bezboruah, T. (2018). A Real-Time Data Acquisition System for Monitoring Sensor Data. Retrieved from http://www.researchgate.net/profile/Pratiksha_Sarma/publication/326398953_A_Real-

Time_Data_Acquisition_System_for_Monitoring_Sensor_Data/links/5b4a05b7aca272c609409d43/A-Real-Time-Data-Acquisition-System-for-Monitoring-Sensor-Data.pdf

Shiel, P. & West, R. (2015). Re-examination of external temperature as a predictor of energy usage in buildings. *Journal of Building Engineering, 3*, 184-194. http://dx.doi.org/10.1016/j.jobe.2015.07.004

Shiel, P. & West, R. (2016). Effects of building energy optimisation on the predictive accuracy of external temperature in forecasting models. *Journal of Building Engineering, 7*, 281-291. http://dx.doi.org/10.1016/j.jobe.2016.07.001

Tian, Z., Si, B., Shi, X. & Fang, Z. (2019). An application of Bayesian Network approach for selecting energy efficient HVAC systems. *Journal of Building Engineering, 25*, 100796. http://dx.doi.org/10.1016/j.jobe.2019.100796

Williams, J., Mitchell, R., Raicic, V., Vellei, M., Mustard, G., Wismayer, A., Yin, X., Davey, S., Shakil, M., Yang, Y., Parkin, A. & Coley, D. (2016). Less is more: A review of low energy standards and the urgent need for an international universal zero energy standard. *Journal of Building Engineering, 6*, 65-74. http://dx.doi.org/10.1016/j.jobe.2016.02.007

Xu, M., Watanachaturaporn, P., Varshney, P. K. & Arora, M. K. (2005). Decision tree regression for soft classification of remote sensing data. *Remote Sensing of Environment, 97*, 322-336. http://dx.doi.org/10.1016/j.rse.2005.05.008

Yuan, C., Liu, S. & Fang, Z. (2016). Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model. *Energy, 100*, 384-390. http://dx.doi.org/10.1016/j.energy.2016.02.001

Zhu, Y., Saeidi, S., Rizzuto, T., Roetzel, A. & Kooima, R. (2018). Potential and challenges of immersive virtual environments for occupant energy behavior modeling and validation: A literature review. *Journal of Building Engineering, 19*, 302-319. http://dx.doi.org/10.1016/j.jobe.2018.05.017