

# Building Energy Management Through Fault Detection Analysis Using Pattern Recognition Techniques Applied on Residual Neural Networks

Imran Khan<sup>1</sup>, Alfonso Capozzoli<sup>1</sup>, Fiorella Lauro<sup>1,3</sup>(✉),  
Stefano Paolo Corgnati<sup>1</sup>, and Stefano Pizzuti<sup>2,3</sup>

<sup>1</sup> Department of Energy (DENERG), Politecnico di Torino, Torino, Italy

<sup>2</sup> Italian National Agency for New Technologies,

Energy and Sustainable Economic Development (ENEA), Rome, Italy

<sup>3</sup> Computer Science and Automation Department, University Roma Tre, Roma, Italy  
fiorella.lauro@polito.it

**Abstract.** In this paper a fault detection analysis through a neural networks ensembling approach and statistical pattern recognition techniques is presented. Abnormal consumption or faults are detected by analyzing the residual values, which are the difference between the expected and the real operating data. The residuals are more sensitive to faults and insensitive to noise. In this study, first, the experimentation is carried out over two months monitoring data set for the lighting energy consumption of an actual office building. Using a fault free data set for the training, an artificial neural networks ensemble (ANNE) is used for the estimation of hourly lighting energy consumption in normal operational conditions. The fault detection is performed through the analysis of the magnitude of residuals using peak outliers detection method. Second, the fault detection analysis is also carried out through statistical pattern recognition techniques on structured residuals of lighting power consumption considering different influencing attributes i.e. number of people, global solar radiation etc. Moreover the results obtained from these methods are compared to minimize the false anomalies and to improve the FDD process. Experimental results show the effectiveness of the ensembling approach in automatic detection of abnormal building lighting energy consumption. The results also indicate that statistical pattern recognition techniques applied to residuals are useful for detecting and isolating the faults as well as noise.

**Keywords:** Building · Energy · ANNs residuals · Pattern recognition · Fault detection

## 1 Introduction

Measuring and collecting large amount of data relevant to building energy consumption and overall performance is becoming increasingly available through

building energy management systems (BEMS). To make the best use of this big data, it is necessary to extract relevant information useful for energy optimization. In this paper different pattern recognition techniques i.e. classification and clustering are used for fault detection.

First, the fault detection is performed by evaluating the magnitude of the residuals generated by an artificial neural network ensemble (ANNE) using an outliers detection method (peak detection). The hourly energy consumption and maximum power for artificial lighting are monitored and used as targets (outputs) of the analyzed models. Second, the fault detection analysis is also performed through statistical pattern recognition techniques (CART, KMeans and DBSCAN) on structured residuals of peak power consumption for lighting considering different influencing attributes i.e. number of people, global solar radiation etc. In this research the capability of ANNE approach and effectiveness of statistical pattern recognition techniques using peak energy consumption residuals for artificial lighting fault detection of a real office building are investigated.

It should be noted that the fault detection of building lighting consumption is not a critical issue that requires a strictly real time execution. Thus, once the ANN models are trained and the ensemble and pattern recognition models are defined, all the simulations can be performed in minute order time. The proposed methodology allow to perform a fault detection analysis in “near” real time, i.e. with a shift of one hour, since an hourly timestamp was considered.

## 2 Motivation and Related Work

Buildings are one of the prime targets to reduce energy consumption around the world. Almost 32% of the total energy consumption in industrialized countries is used for electricity, heating, ventilation, and air conditioning (HVAC) in buildings [8]. Furthermore, building industry is not only energy-intensive, but also knowledge-intensive. The real data of a building contains the actual information of building operation; and thus can reflect the building performance accurately [22]. For energy optimization, the evaluation of real time building energy consumption data is a demandable and emerging area of building energy analysis. Several studies have been published on methods for automatically detecting abnormal energy consumption data in buildings. Seem [19] presented pattern recognition with robust statistical outliers’ detection method to investigate abnormal energy consumption. Liu et al. [12] used classification and regression tree method for whole building energy abnormal behavior. Some research works, [11, 18, 21], provided classification methods including the box plots approach, association rule mining and pattern recognition algorithm to detect anomalous energy consumption in buildings.

Yu et al. [22] used fuzzy neural networks model for fault detection and diagnosis on the energy consumption of the whole building. Fault-free measured data is used to build up the model and another measured data with a fault is used to validate the model and test the performance of fault detection. Model is applied on the measured data with fault of an open window in the room, and threshold

for the fault detection is derived from the moving mean value and variance in a certain period (24 h). Dodier and Kreider [3] proposed neural network algorithm to evaluate whether the energy consumption data is normal. The energy consumption is predicted by collection of previous data by neural network. The ratio of actual energy consumption to expected energy consumption is calculated. The data is considered abnormal when the ratio is lower or higher than thresholds. Also, Holcomb et al. [6] proposed algorithmic techniques based on machine learning to address the prediction of building energy consumption from that of similar buildings in its geographical neighborhood and to localize the faults in building sub-systems.

### 3 Case Study and Data Introduction

The case study selected for the fault detection analysis is an office building located in Rome, Italy. The building is composed of three floors and is equipped with a monitoring system aimed at collecting energy consumption (electrical and thermal) and the environmental conditions. Moreover each room/office in the building is equipped with a presence sensor. In the paper, experiments are performed on a data set referred to energy consumption for artificial lighting only for the first floor. In this floor there are 13 offices and two CED rooms. Different number of fluorescent lamps (each 55 W) ranging from 4 to 8 are installed in each office/room. In the two CED rooms 12 lamps, each 55 W, are installed. In order to identify abnormal lighting energy consumption, the features considered as dependent variables for the models are the average hourly energy consumption and peak demand (maximum power). Both lighting energy and power consumption of buildings first floor are analyzed for the months of December and January. Furthermore, the independent variables that are recorded with an hourly timestamp are: people presence, number of active rooms (a room is considered active if at least one person is present), global solar radiation, time, date and day of the week. In order to verify the reliability and the effectiveness of the proposed methods two artificial faults have been created on 24th and 25th of January. In these days at the end of the working time with fewer people presence between 17:30 and 18:00 all artificial lights of the offices on the first floor have been switched on creating a peak of energy demand.

## 4 Brief Description of Proposed Methods

In this section a brief theoretical description of proposed methods is presented.

### 4.1 Consumption Modelling by Artificial Neural Network Basic Ensembling Method

Artificial neural networks (ANNs) [1,5] are black-box (or data-driven) models mainly used when analytical or transparent models cannot be applied to model

complex relationships between inputs and outputs. The basic processing units of ANNs are neurons: the connections between neurons define the network topology or architecture. Among all the different types of interconnecting structures, the feedforward one is widely used: the data processing can extend over multiple (layers of) units, but no feedback connections are present, i.e., connections extending from outputs of units to inputs of units in the same layer or previous layers. These models are also known as multi-layer perceptrons (MLP) [17], since the basic structure is the perceptron [16].

An “ensemble” is a group of learning models working together on the same task to improve the performances of the constituent models. In the last years, several ensembling methods have been carried out [10, 13]. The non-generative ensembling method seeks to combine the outputs of the models in the best way. In the case of ANNs, they are trained on the same data, they run together and their outputs are combined in a single one. In particular, basic ensemble method (BEM) [2, 15], is the simplest non-generative ensembling method: it combines the outputs of  $M$  neural networks as their arithmetic mean.

## 4.2 Peak Detection Method and Mzscore

In many applications, such as building energy consumption analysis and savings, defining “peaks” in an objective way is very important for an easier identification in a given time-series. Thus, a peak can be defined as an observation that is inconsistent with the majority of observations of a data set.

The method considered in this work, peak detection method, calculates the value (score) of a peak function  $S$  for every element of the given time-series [14]. A given point is a peak if its score is positive and it is greater than or equal to a particular threshold value. Particularly, a peak function  $S$  computes the average of the maximum among the signed distances of a given point  $x_i$  in a time-series  $T$  from its  $k$  left neighbours and the maximum among the signed distances from its  $k$  right neighbours. The function  $S$  is an index that allows quantifying the severity of outliers and then provides information about the priorities for actions to be associated with each outlier. In addition to the function  $S$ , another synthetic index modified zscore (*Mzscore*) is used to determine the amount of variation from normal observations. This index is based on the distance and direction of each outlier compared to the average value of normal observations (observations that do not contain outliers).

## 4.3 Classification and Regression Tree (CART)

The CART algorithm is based on classification and regression trees. A CART is a binary decision tree that is constructed by splitting a parent node into two child nodes repeatedly, beginning with the root node that contains the whole learning sample. CART can easily handle both numerical and categorical variables and useful in robust detection of outliers. A decision tree is constructed from the recorded data which can easily be converted to classification rules for effective identification of anomalies. Therefore it is particularly suitable for conducting

analysis of fault detection in real time. CART methodology generally consists of three parts: construction of maximum tree, choice of the right size tree and classification of new data [20].

#### 4.4 Clustering

The selected algorithms can be classified into two categories: (i) partitioning methods and (ii) density-based methods. These methods require the definition of a metric to compute distances between objects in the dataset. In the case study analyzed, distances between objects are measured by means of the Euclidean distance computed on normalized data.

**KMeans.** It belongs to partitioning category [7], is able to find spherical-shaped clusters and is sensitive to the presence of outliers. It requires as input parameter  $k$ , the number of partitions in which the dataset should be divided. It represents each cluster with the mean value of the objects it aggregates, called centroid. The algorithm is based on an iterative procedure, preceded by a set-up phase, where  $k$  objects of the dataset are randomly chosen as the initial centroids. Each iteration performs two steps; in the first step, each object is assigned to the cluster whose centroid is the nearest to that object. In the second step centroids are relocated, by computing the mean of the objects within each cluster. Iterations continue until the  $k$  centroids do not change.

**DBSCAN.** It is a density-based method designed to deal with non-spherical shaped clusters and is less sensitive to the presence of outliers. DBSCAN [4] requires two input parameters, a real number  $r$  and an integer number  $minPts$ , used to define a density threshold in the data space. A high density area in the data space is an  $n$ -dimensional sphere with radius  $r$  which contains at least  $minPts$  objects. DBSCAN is an iterative algorithm which iterates over the objects in the dataset, analyzing their neighborhood. The effectiveness of the algorithm is strongly affected by the setting of parameters  $r$  and  $minPts$ .

## 5 Results and Analysis

The ANN ensemble is built according to BEM, considering 10 feed-forward MLP ANNs, with 1 hidden layer consisting of 15 neurons, hyperbolic tangent as activation function for the hidden neurons, and linear for the output. The training period is approximately 4 weeks and testing period approximately 1 week. Simulations are performed with MATLAB R2012b through the Levenberg-Marquardt algorithm. The reported results (see Table 1) are averaged over the 10 different runs (standard deviation in brackets). Performance has been evaluated according to the mean absolute error (MAE) and the maximum absolute error (MAX):

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

**Table 1.** Experimental results (training and testing)

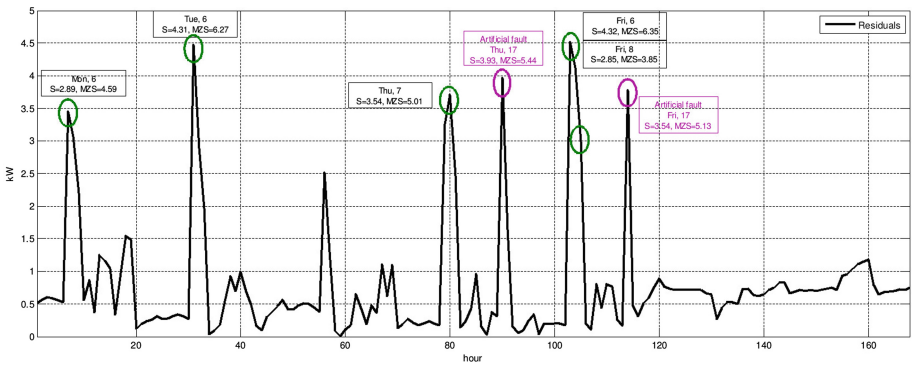
	TRAINING	ANN	BEM	TESTING	ANN	BEM
Active energy	MAE (kWh)	0.33( $\pm 0.01$ )	0.31	MAE (kWh)	0.66( $\pm 0.04$ )	0.63
	MAX (kWh)	1.41	1.06	MAX (kWh)	4.26	3.75
Maximum active power	MAE (kW)	0.35( $\pm 0.02$ )	0.33	MAE (kW)	0.81( $\pm 0.05$ )	0.78
	MAX (kW)	1.76	1.4	MAX (kW)	4.78	4.51

$$MAX = \max \{|y_i - \hat{y}_i|\}_{i=1}^N \quad (2)$$

where  $y_i$  is the real lighting consumption,  $\hat{y}_i$  is the output of the model (estimated lighting consumption) and  $N$  is the size of the real data set.

As shown in Table 1, the results obtained with ANN BEM are slightly better than those obtained with constituent ANNs. In the following sections only the analysis performed on the maximum power for lighting is presented.

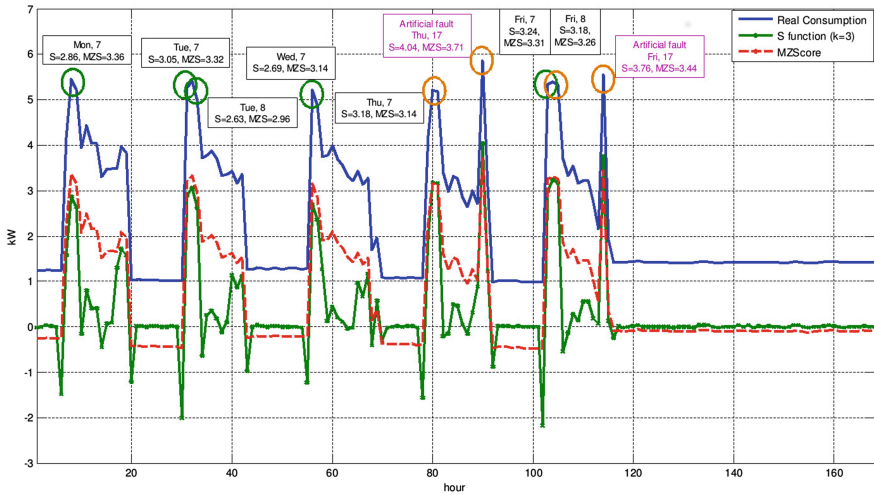
In order to estimate a normal pattern of the maximum electrical power for the artificial lighting, the training of the ANN BEM is performed considering a fault free data set, obtained through outlier detection. The lighting power demand is estimated really well through the ANN BEM in the training period. In the testing period the estimated power follow quite well the monitored power demand, with the exception of some evident abnormal values. The magnitude of the difference over the time between the actual and estimated power demand is analyzed for detecting anomalous situations. To this purpose the peak detection method has been applied to the residuals data set in the testing period. In Fig. 1 the trend of residuals over the time is shown and the abnormal detected power demand values are highlighted. The identified residual peaks include potential early morning faults, for which very high power demand is observed corresponding to the only cleaning staffs presence, and the two artificial faults. The results confirm that the analysis of residuals generated through the ANN BEM represents a useful and powerful technique for the peak building lighting fault detection.

**Fig. 1.** Testing residuals (maximum active power) and detected peaks

Then, the peak detection method is applied to the maximum power consumption time-series. In Fig. 2 the outliers detected for testing period are shown with the relative values of Mzscore and S function indices. It can be observed that the method allows detecting the two artificial faults and some other real faults in early morning. In these situations the relative severity indices correctly assume higher value. However, the data show that power is related to other variables i.e. people, solar radiation, day and active rooms, so it can be inferred that the extreme values are not always definite faults. Therefore some false positives can be found when a univariate outlier detection method is applied without taking into account the effect of the independent variables on the consumptions.

In second part of this research, statistical pattern recognition techniques are applied on structured residuals of lighting power consumption. The major steps adopted for this analysis of fault detection are summarized as:

- Sensitivity analysis is carried out to identify the independent variable(s) of greater importance on the variation of the dependent variable (maximum power residuals), as shown in Fig. 3.
- CART is used for classification with one pruning method (number of cases in parent and child nodes). After multiple simulations and thorough analysis of the constructed classes, the number of cases set for parent and child nodes are 40 and 20 respectively. The independent variables considered for the classification are day, date, time, people presence, active rooms and solar radiations. The data are divided into 4 classes and each class has been analyzed separately. The classes are formed with time as most influential factor which is also evident from Fig. 3. In Fig. 4a and b scatter plots for class 1 are shown highlighting the two artificial faults. The class 1 contains the data values of early



**Fig. 2.** Maximum active power (testing period), S function values, mzscores and detected peaks (common peaks are orange) (Color figure online)

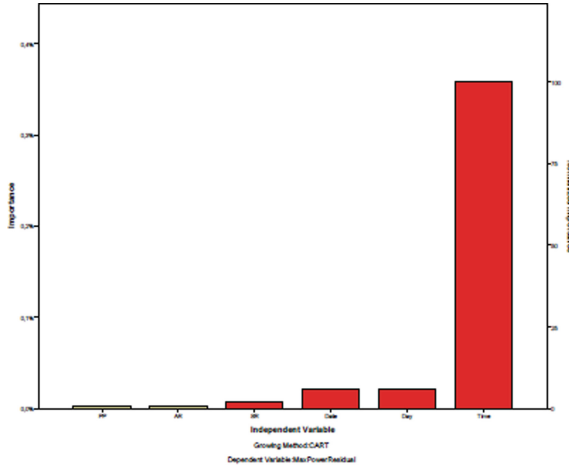


Fig. 3. Sensitivity analysis

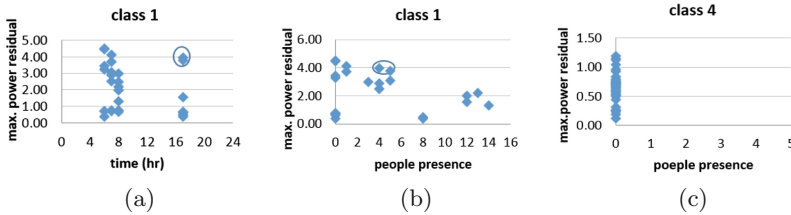


Fig. 4. Scatter plots for class 1 (CART) with artificial outliers encircled and for class 4

morning (06:00–08:00) and evening (17:00) for all week. Thorough analysis of class 1 shows that the most values are outliers including two artificial faults as with fewer numbers of people and/or active rooms the electrical power consumption for lighting is high. Classes 4, 5 and 6 are mostly pure and do not contain abnormal values. From scatter plot of class 4 (see Fig. 4c), it can be seen that the class contains zero number of people and most values in that class are normal. Also the number of active rooms is always zero and the values of global solar radiation are mostly zero or fewer in that class. Class 6 mostly consists of high number of active rooms and people (10 or more) and high values of solar radiation.

- For clustering (KMeans and DBSCAN), in order to overcome the limitations of the algorithms that do not allow time and day as independent variables, data sets have been divided into the working period (07:00–18:00), the non-working period and weekends. The approach adopted for the splitting of the data is the experience gained from our previous work [9] for which the division of the data set in the daytime, nighttime and weekend proved not to be particularly effective for the nature of the fault in the type of building under investigation.



It did not allow effectively the detection of outliers present in the early hours of the morning and at the end of working hours. Also before performing the clustering analysis, values of both dependent (max. power residuals) and independent variables (people presence, active rooms and solar radiations) have been normalized by means of standard score (z-score) method. For working, non-working and weekends the data is divided into 3, 2 and 2 clusters respectively using KMeans clustering. For DBSCAN clustering all three different split data sets are divided into two clusters each. To set the input parameters ( $r$ ,  $minPts$ ) multiple tests are carried out for each data set by using different values for these parameters. With DBSCAN method, in all discovered clusters, the cluster label zero contains all points identified as outliers or noise.

Tables 2 and 3 show the cluster 1 (KMeans) and cluster zero (DBSCAN) clustering respectively performed on working hours data set. The results show that the splitting criteria used in this research has proved effective in overcoming the limitations of cluster algorithms and produced good results. Both the clusters are impure and contain artificial outliers with other positive outliers too. Clusters 2 and 3 (KMeans-working hours) are pure. In Fig. 5a and b scatter plots for cluster 2 (KMeans) are given. The cluster includes higher number of people and active rooms and energy consumption can be considered as normal.

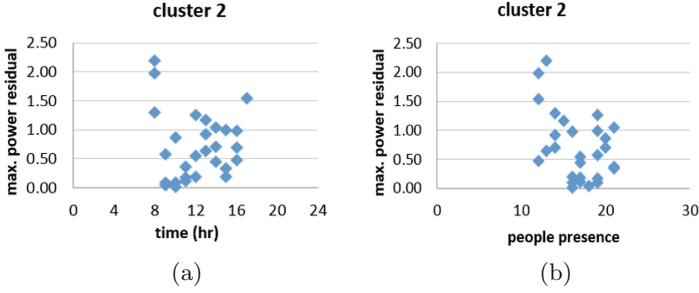
For non-working hours, both cluster 2 (KMeans) and zero cluster (DBSCAN) have higher values of energy consumption corresponding to zero number of people presence in the early morning (06:00 am). For weekends, the clusters are formed

**Table 2.** KMeans (working hours, cluster 1), artificial faults are highlighted

Day	Date	Hour	Maximum power residual	Maximum power (kW)	PP	AR	SR	Clu-1
Mon	21/01/2013	7	3.07	5.45	5	5	0.83	1
Tue	22/01/2013	7	2.88	5.41	4	4	3.80	1
Wed	23/01/2013	7	2.51	5.2	4	4	7.00	1
Thu	24/01/2013	7	3.70	5.2	1	1	2.00	1
Thu	24/01/2013	8	2.47	5.19	4	3	27.00	1
<b>Thu</b>	<b>24/01/2013</b>	<b>17</b>	<b>3.96</b>	<b>5.86</b>	<b>4</b>	<b>3</b>	<b>2.67</b>	<b>1</b>
Thu	24/01/2013	18	1.64	3.08	3	2	0.00	1
Fri	25/01/2013	7	4.12	5.39	1	1	2.50	1
Fri	25/01/2013	8	3.00	5.33	3	3	19.62	1
<b>Fri</b>	<b>25/01/2013</b>	<b>17</b>	<b>3.77</b>	<b>5.55</b>	<b>5</b>	<b>4</b>	<b>1.75</b>	<b>1</b>

**Table 3.** DBSCAN (working hours, cluster 0), artificial faults are highlighted

Day	Date	Hour	Maximum power residual	Maximum power (kW)	PP	AR	SR	DBSCAN
Thu	24/01/2013	7	3.70	5.20	1	1	2.00	cluster_0
Thu	24/01/2013	12	0.96	3.27	6	5	476.45	cluster_0
<b>Thu</b>	<b>24/01/2013</b>	<b>17</b>	<b>3.96</b>	<b>5.86</b>	<b>4</b>	<b>3</b>	<b>2.67</b>	<b>cluster_0</b>
Fri	25/01/2013	7	4.12	5.39	1	1	2.50	cluster_0
Fri	25/01/2013	8	3.00	5.33	3	3	19.62	cluster_0
<b>Fri</b>	<b>25/01/2013</b>	<b>17</b>	<b>3.77</b>	<b>5.55</b>	<b>5</b>	<b>4</b>	<b>1.75</b>	<b>cluster_0</b>



**Fig. 5.** Scatter plots for cluster 2 (KMeans-working hours)

**Table 4.** Some common outliers in all methods

Day	Date	Hour	Maximum power residual	Maximum power (kW)	PP	AR	SR
Mon	21/01/2013	6	3.45	4.17	0	0	0.00
Mon	21/01/2013	7	3.07	5.45	5	5	0.83
Mon	21/01/2013	8	2.20	5.23	13	9	19.75
Tue	22/01/2013	6	4.46	5.26	0	0	0.00
Tue	22/01/2013	7	2.88	5.41	4	4	3.80
Wed	23/01/2013	7	2.51	5.2	4	4	7.00
Thu	24/01/2013	6	3.26	4.2	0	0	0.00
Thu	24/01/2013	7	3.70	5.2	1	1	2.00
Thu	24/01/2013	8	2.47	5.19	4	3	27.00
<b>Thu</b>	<b>24/01/2013</b>	<b>17</b>	<b>3.96</b>	<b>5.86</b>	<b>4</b>	<b>3</b>	<b>2.67</b>
Fri	25/01/2013	6	4.51	5.34	0	0	0.00
Fri	25/01/2013	7	4.12	5.39	1	1	2.50
Fri	25/01/2013	8	3.00	5.33	3	3	19.62
<b>Fri</b>	<b>25/01/2013</b>	<b>17</b>	<b>3.77</b>	<b>5.55</b>	<b>5</b>	<b>4</b>	<b>1.75</b>

with solar radiation as the most influential factor. Also the outliers detected by each method are compared and some common outliers are presented in Table 4. By analyzing the results obtained from each method, it can be concluded that, in general, outliers are identified in two different periods of the day. The first period is early morning (06:00–08:00 am). In the early morning electrical power for lighting has the peaks at a very low presence of occupants. The second period is related to the end of working hours (17:00), where it is observed that a decrease in the number of occupants of the building does not correspond to a decrease in electrical power consumption for lighting.

## 6 Conclusions

To achieve objectives for energy efficiency, it is necessary to evaluate information contained in sensed building data. This paper presented a new approach by combining the ANNE with statistical pattern recognition techniques for evaluating

real energy consumption data for whole building lighting energy fault detection. The research is aimed at investigating the potential of using ensembling techniques and, additionally, usefulness of statistical pattern recognition methods performed on structured residuals for fault detection.

The fault detection, performed through the analysis of the magnitude of residuals using a peak detection method, allowed to detect the two artificial faults and some other actual anomalous power values in the testing data set. Finally, the results obtained through all statistical pattern recognition techniques, performed on structured residuals, proved to be adequate and each method has been able to detect artificial faults and other positive outliers.

The application of this approach can improve fault detection process by reducing the number of false anomalies. The data set considered in this study is relatively small and only for artificial lighting of single building. In future, the work will consider new end-uses i.e. HVAC, plug load and exploit other data mining techniques for fault detection.

## References

1. Arbib, M.A.: *The Handbook of Brain Theory and Neural Networks*. MIT press, Cambridge (2003)
2. Bishop, C.M.: *Neural Networks for Pattern Recognition*. Oxford University Press Inc, New York (1995)
3. Dodier, R.H., Kreider, J.F.: Symposium papers-ch-99-5-fault detection and diagnostics-learning from building operations-detecting whole building energy problems. *ASHRAE Trans. Am. Soc. Heating Refrig. Airconditioning Engin* **105**(1), 579–592 (1999)
4. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *KDD*, vol. 96, pp. 226–231 (1996)
5. Haykin, S.: *Neural Networks: A Comprehensive Foundation*, 2nd edn. Prentice Hall PTR, Upper Saddle River (1998)
6. Holcomb, D., Li, W., Seshia, S.A.: Algorithms for green buildings: learning-based techniques for energy prediction and fault diagnosis. *Google Scholar, UCB/EECS-2009-138* (2009)
7. Juang, B.H., Rabiner, L.: The segmental k-means algorithm for estimating parameters of hidden markov models. *IEEE Trans. Acoust. Speech Sig. Process.* **38**(9), 1639–1641 (1990)
8. Katipamula, S., Brambley, M.R.: Review article: methods for fault detection, diagnostics, and prognostics for building systems a review, part i. *HVAC&R Res.* **11**(1), 3–25 (2005)
9. Khan, I., Capozzoli, A., Corgnati, S.P., Cerquitelli, T.: Fault detection analysis of building energy consumption using data mining techniques. *Energy Procedia* **42**, 557–566 (2013)
10. Krogh, A., Vedelsby, J., et al.: Neural network ensembles, cross validation, and active learning. In: Tesauro, G. (ed.) *Advances in Neural Information Processing Systems*, pp. 231–238. MIT Press, Cambridge (1995)
11. Li, X., Bowers, C.P., Schnier, T.: Classification of energy consumption in buildings with outlier detection. *IEEE Trans. Ind. Electron.* **57**(11), 3639–3644 (2010)

12. Liu, D., Chen, Q., Mori, K., Kida, Y.: A method for detecting abnormal electricity energy consumption in buildings. *J. Comput. Inf. Syst.* **6**(14), 4887–4895 (2010)
13. Liu, Y., Yao, X.: Ensemble learning via negative correlation. *Neural Netw.* **12**(10), 1399–1404 (1999)
14. Palshikar, G., et al.: Simple algorithms for peak detection in time-series. In: *Proceedings of the 1st International Conference Advanced Data Analysis, Business Analytics and Intelligence* (2009)
15. Perrone, M.P., Cooper, L.N.: When networks disagree: ensemble methods for hybrid neural networks. Technical report, DTIC Document (1992)
16. Rosenblatt, F.: The perceptron—a perceiving and recognizing automaton. Technical report 85–460-1, Cornell Aeronautical Laboratory, Buffalo, NY, 00344 (1957)
17. Rosenblatt, F.: Principles of neurodynamics. Perceptrons and the theory of brain mechanisms. Technical report, DTIC Document (1961)
18. Seem, J.E.: Pattern recognition algorithm for determining days of the week with similar energy consumption profiles. *Energy Buildings* **37**(2), 127–139 (2005). <http://www.sciencedirect.com/science/article/pii/S0378778804001434>
19. Seem, J.E.: Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy Buildings* **39**(1), 52–58 (2007). <http://www.sciencedirect.com/science/article/pii/S0378778806001514HrB>
20. Timofeev, R.: Classification and regression trees (cart) theory and applications (2004)
21. Yoshida, K., Inui, M., Yairi, T., Machida, K., Shioya, M., Masukawa, Y.: Identification of causal variables for building energy fault detection by semi-supervised lda and decision boundary analysis. In: *2008 IEEE International Conference on Data Mining Workshops, ICDMW'08*, pp. 164–173. IEEE (2008)
22. Yu, B., van Paassen, D.H.: Fuzzy neural networks model for building energy diagnosis. In: *8th International IBPSA Conference*, pp. 1459–1466 (2003)