Energy demand prediction in smart buildings using advanced machine learning techniques

Desiree Arias-Requejo[1 2 3] Carlos J. Alonso-Gonzalez[4], Belarmino Pulido[4], Marcus M. Keane[1 2 3]
[1]School of Engineering, College of Science and Engineering, National University of Ireland Galway, Ireland. desiree.arias@nuigalway.ie*. marcus.keane@nuigalway.ie
[2]Informatics Research Unit for Sustainable Engineering (IRUSE) Galway, Ireland
[3]Ryan Institute, National University of Ireland Galway, Ireland
[4]Grupo de Sistemas Inteligentes, Departamento de Informática, Universidad de Valladolid, Spain. calonso@infor.uva.es, belar@infor.uva.es

Abstract

Reduction of energy consumption is essential to reduce energy waste. This is even more important in the building sector that accounts for 27% of total CO_2 emissions in Europe. In this work, we propose to use available data from a smart building in the NUIG campus (Ireland) to generate black-box models for energy demand prediction using advanced machine-learning techniques. In this paper, we present the first step for an accurate estimation of the energy consumption in buildings. Firstly, hierarchical clustering is used to find the most probable system health state. Secondly, the energy demand models for that state are used to estimate the intended energy consumption. Experimental results showed that the energy demand models specific for a state performs better than a general model for all the system states, confirming our initial hypothesis. This work can be the foundation to perform predictive maintenance based on the energy prediction. In the absence of system faults, a deviation in the energy consumption can be related to tear and wear problems, thus prompting the need for maintenance. Consequently, the reduction in energy consumption due to early detection of a degradation problem will also help to reduce maintenance costs.

Introduction

Building sector is one of the greatest energy consumers, accounting for 40% of worldwide primary energy and more than 25% in Europe (Costa et al., 2013; International Energy Agency, 2018b). It is also the second largest emitter of CO_2 with 27% of total CO_2 emissions in Europe (International Energy Agency, 2018a). Buildings in the United Kingdom are estimated to have an avoidable energy waste due to inefficient operation conditions between 20 to 50% (Warburton et al., 2009). In buildings, heating, ventilation, and air conditioning (HVAC) systems are one of the major energy consumers. Moreover, one of the reasons why these systems are often inefficient is the presence of undetected failures (Sterling et al., 2014). Air Handling Units (AHUs) are the mechanical ventilation systems of the HVAC systems in charge of air conditioning and circulation.

Energy use improvement in buildings has been an active area in research. A large variety of methods and approaches has been proposed to improve the energy use in buildings. Model-based or physic-based methods are based on analytical and mathematical models and usually have high accuracy and flexibility in the configuration. However, to develop physics-based (white-box) models for large and complex buildings and systems can be expensive and complex (Naug et al., 2020). An alternative to these models, is data-driven or black-box models. They required a large quantity of historical data to learn the behaviour of building systems (Naug et al., 2020). Data-driven methods can be grouped by their type in benchmarking models, energy-mapping models, energy forecasting models and energy profiling models (Ahmad et al., 2018). Finally, hybrid approaches are a combination of the two previous approaches.

New and more restrictive policies for energy efficiency in buildings are being applied (Medojevic et al., 2018). The advances in IoT devices, Internet connection speeds and processing capabilities of computers have enabled more and more buildings to be equipped with sensors/meters that generate huge amounts of data every day. This fact enables that they can be monitored continuously from an environmental and energy performance perspective, and thus favouring the use of data-driven techniques.

In the area of machine learning, deep learning techniques are gaining attention and are being applied in different areas thanks to its great performance. Short-term energy forecasting with data-driven techniques has been the subject of several research publications. A review of Recurrent Neural Networks (RNN) for short-term load forecasting was carried out by Bianchi et al., (2017). Deep learning methods have been also used to estimate short-term electric load in smart grids (Gasparin et al., 2019). Models based on Restricted Boltzmann Machines (RBMs), a deep learning method, outperformed other methods like Artificial Neural Networks (ANN), Support Vector Machine (SVM) and RNN to forecast the energy consumption in a residential building (Mocanu et al., 2016). A special artificial RNN, Long-short Term Memory (LSTM) algorithms (Hochreiter & Schmidhuber, 1997) have been widely used to solve sequence learning problems, especially in the area of natural language processing (Graves et al., 2013; Habernal &

Matoušek, 2013). Recently, these algorithms have begun to be applied to energy forecasting. Architectures based on LSTM algorithms, have shown good performance in the energy forecasting on a benchmark dataset (Marino et al., 2016). A study on the forecasting of energy consumption of air-conditioning systems, compared the performance of LSTM models with two more traditional models namely back propagation (BP) neural networks and Autoregressive Integrated Moving Average (ARIMA) timeseries model (Zhou et al., 2020). In that study, experimental results showed that LSTM obtained better results in both hourly and daily predictions.

The combination of clustering techniques and learning algorithms such as SVM or Adaptative Boosting (AdaBoost) obtained successfully results in the prediction of the energy consumption of a three-storey building and potential energy savings were provided (Naug & Biswas, 2018).

In our research we propose a data-driven approach for building energy forecasting. As buildings systems operate in different modes during seasons and day times, a unique model could not be enough to characterise the system behaviour (Naug & Biswas, 2018). Our hypotheses are that unsupervised learning, i.e. clustering, can help to characterise the different operational models, and then using deep-learning models, we can develop a specific model for each cluster. We will test our proposal using real data collected from a weather station and from sensors installed on a smart building. Our initial guess is that by applying clustering to weather data, different environmental conditions that can represent those operating modes can be identified. A specific model for each operation mode provides more accurate predictions than a global energy model. Thanks to an accurate energy model, system health can be monitored. Using those predictive models, deviations in the energy consumption related to tear and wear problems can be spotted in advance and maintenance actions can be scheduled before a system fault occurs.

In the following section we provide a description of the proposed methodology to address the problem of energy forecasting in buildings under different operational modes. Next, we introduce the case study and present and discuss the results obtained and we conclude summarising the work done and presenting the future work.

Methodology

Our proposed methodology is based on the assumption that different environmental conditions influence the operation of the building and its HVAC systems, thus obtaining different energy demand models. Knowing the kind of environmental conditions, specific models for energy forecasting can be applied. To prove this assumption, we compared the performance between specific energy forecasting models for a cluster and general energy forecasting models. This general energy forecasting models are trained to forecast the energy consumption regardless of the system operational mode. To identify the different operational modes, we will use two unsupervised algorithms, k-means and hierarchical clustering. Finally, for the energy forecasting models we will use LSTM algorithms. LSTM models are specialised in learning and storing information over large intervals of time and were designed to overcome the vanishing gradient problems that usually appear when using RNN (Hochreiter & Schmidhuber, 1997).

Case study

The Alice Perry building belongs to the National University of Ireland at Galway and was opened in 2011. This four-storey building has 400 rooms and covers 14,250 square meters. Thanks to its multiple and diverse sensors (approximately 4,000), it serves as a living laboratory providing multiple live data sets. It has a strong focus on green-building initiatives to reduce its energy waste. In this work, we have used the total thermal energy demand for the heating coils of the eleven AHUs of the Alice Perry.

Although we have information about internal conditions of the rooms in the building, our guess is that external environmental conditions should be much more meaningful to identify different working points or operation modes in the HVAC system. The weather data used in this work, has been collected by the weather station installed on the roof of one of the buildings of the NUIG campus.

Dataset description

The "weather" dataset contains the data collected by the weather station between 2018 and 2019. There are two periods at the beginning of 2019 without data. The available set of measurements in this dataset are (in round brackets the abbreviations):

- Panel temperature in Celsius degrees.
- Dry bulb air temperature (temp) in Celsius degrees.
- Relative humidity (RH) in percentage.
- Total solar irradiation in kW/m^2 (Sir [kW/m2] Tot).
- Total solar irradiation in kJ/m^2 (solarIrr).
- Wind direction (winddir) in degrees.
- Average of one minute of the wind speed (windspeed) in m/s (Wind Speed [m/s]).
- Maximum three second gust over a one-minute period in m/s (Gust 3s Avg [m/s]).
- Rainfall (rain) in mm.
- Barometric pressure in mBar.

During the data exploration, we identified, in the form of huge outliers, some incorrect measurements in the barometric pressure. Those incorrect measurements indicated a barometric pressure of 600 mBar, while the rest of the measurements were between 970 and 1050 mBar. Those incorrect instances were deleted. Using box-plots we also noticed that during spring and summer there was more variation in the air temperature, the solar irradiation and the relative humidity, while the barometric pressure had less variation in those seasons. July is the month with higher median air temperature. The typical solar irradiation during winter is very low, while the wind speed has approximately the same distribution during the year. There are a lot of outliers in the rain attribute, since most measurements are zero. Figure 1 shows the evolution of the variable RH during 2018.

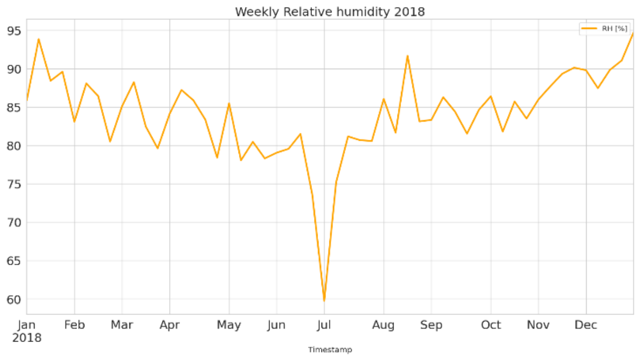

Figure 1. Weekly mean of the relative humidity in 2018.

Clustering

We used crisp partitioning, in which, each instance can only belong to one specific cluster. The unsupervised techniques to obtain the clusters chosen were k-means and hierarchical clustering.

The unsupervised algorithm k-means generates minimum-squared-error clusters (Pakhira et al., 2004). Usually, the optimal number of initial centroids is unknown. Therefore, we applied the Elbow Method in order to determine the ideal quantity of centroids based on the WCSS (Within Clusters Summed Squares) optimization. In this way, we calculated the sum of squared distances from each instance to its assigned centroid. Then, we determined the ideal number of centroids getting the point where the variations in the WCSS start to become not-significant. We do not include the graphics due to lack of space, but the selected numbers of centroids were five, four and three.

Agglomerative or bottom-up approach was the method we selected to construct the hierarchical clustering. In this method, clusters are iteratively merged using a similarity measure until reached the selected number of clusters (Rokach & Maimon, 2005). As similarity measure, we selected the minimum variance method (Ward) with the Euclidean metric. One of the advantages the hierarchical clustering offers is the possibility of construct a dendrogram that can ease the decision on the number of clusters.

Data used for the clustering

As mentioned in the dataset description, there are 10 weather related variables. Some of them were discarded since they were correlated with other attributes. For instance, the "panel temperature" was highly correlated with "air temperature", so we can omit the first one. Also, we can omit the variable "Gust 3s Avg" as it is correlated with Wind Speed. Finally, we also omit "Sir Tot" as it is correlated with "Sir [kW/m2]".

Consequently, the set of selected weather variables for the clustering problem are: dry-bulb temperature, relative humidity, barometric pressure, wind speed, wind direction and total and diffuse solar irradiance, and rainfall. Except for rainfall (which is measured hourly), all the data were sampled every minute. and hourly data of rainfall. These are the attributes we will use in our clustering. The weather data used was resampled to the weekly mean, obtaining 52 instances.

For the clustering, we used the data collected from the weather station in 2018.

Feature extraction

To select the relevant weather variables for the clustering, we have applied Independent Component Analysis (ICA). ICA algorithms can be used for blind source separation and feature extraction. There are different implementations of ICA and the independent components (ICs) extracted depend on which algorithm is selected (Ozawa & Kotani, 2000). Fast ICA algorithm (Hyvarinen & Oja, 1997) implemented in scikit-learn (Pedregosa et al., 2011) was the algorithm selected due to its fast convergence to the most accurate solution. In this step, we applied the linear transformation called *whitening* to remove any existent correlation in the data. Being v the dimension or number of original features we had, we obtained a v-dimensional ICA feature vectors. We did the feature selection for this ICs vectors based on Kurtosis. Considering that the most useful features are the most independent, we ranked the features according to its kurtosis absolute value and we selected the greatest ones (Ozawa & Kotani, 2000). Accordingly, we selected the four ICs with the greatest non-gaussian distribution. However, after the clustering evaluation, we obtained better results with only three ICs.

Clustering construction

In the initial steps of the cluster construction, we observed, as Figure 2 shows, that there was a cluster that was only composed by one instance (instance 25), which corresponded to the last week of June. In that week, we observed that the relative humidity dropped drastically (see Figure 1), and both the solar total irradiation and the air temperature had uncommonly high values. As that cluster was isolated, we drop that instance and create another dataset for the clustering without that instance, to test if we obtain better clusters with that dataset (see Figure 3). We will refer to this new dataset as "without outlier" in the rest of the paper.

Several experiments were carried out to determine the best combination of attributes and number of clusters. We compared different subsets of original weather attributes with the ICs obtained. Being k the targeted number of clusters, for each clustering algorithm, k-means and hierarchical clustering, we did the clustering stablishing k

with values 5 and 4 for the original dataset, and with values 4 and 3 for the dataset without outlier. The first set of attributes tested was air temperature, relative humidity, and solar irradiation (referred from now on as “3 attributes”). Then, we added wind speed (referred as “4 attributes”) and finally we also introduced the rainfall attribute (“5 attributes”). We also did the clustering using first the three most significant ICs and later with the four most significant ICs. In total we did 40 different experiments (2 algorithms x 4 number of clusters x 5 attributes sets).

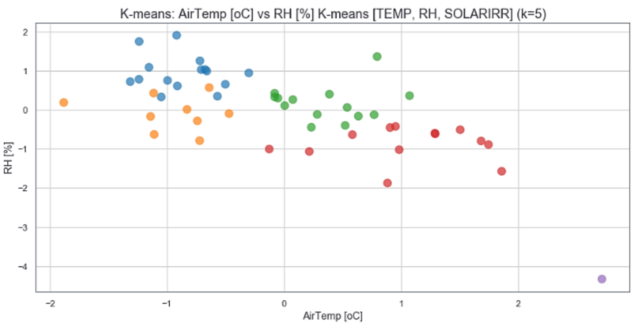

Figure 2. Example of k-means clustering using the original dataset with k=5 using the attributes: air temperature, relative humidity, and solar irradiation.

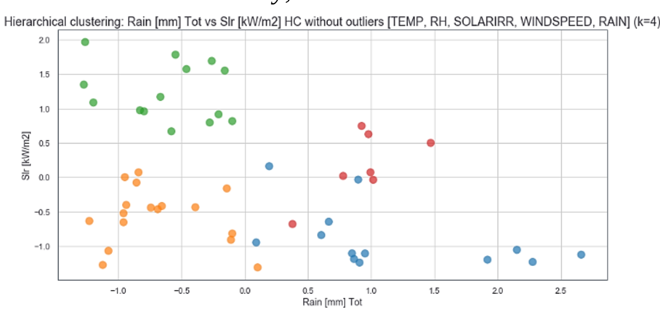

Figure 3. Example of hierarchical clustering of the data without outlier with k=4 using the attributes: air temperature, relative humidity, solar irradiation, wind speed and rainfall.

Clustering performance evaluation

To evaluate the compactness and separation of the clusters obtained in the 40 experiments, we considered three metrics: Silhouette (SIL) (Rousseeuw, 1987), Davies-Bouldin (DB) (Davies & Bouldin, 1979) and PBM (Pakhira et al., 2004). Silhouette and Davies-Bouldin metric were already implemented in scikit-learn, but PBM was not, so we implemented it. The objective using PBM index is to maximise it. PBM can be defined as (Pakhira et al., 2004):

$$PBM(K) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K\right)^2 \qquad (1)$$

In which:

$$E_K = \sum_{k=1}^{K} E_k \qquad (2)$$

$$E_k = \sum_{j=1}^{n} u_{kj} \left| x_j - z_k \right| \qquad (3)$$

$$D_K = \max_{1 \le i,j \ge K} \left| z_i - z_j \right| \qquad (4)$$

$$U(X) = \left[u_{kj} \right]_{K \times n} \qquad (5)$$

Being K the number of clusters, n the number of points in the dataset, z_k the centroid of the k^{th} cluster and $U(X)$ the partition matrix for the data.

Silhouette score $s(i)$ for a single sample i can be defined as (Rousseeuw, 1987):

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \qquad (6)$$

Where $a(i)$ is the mean intra-cluster distance and $b(i)$ is the mean nearest-cluster distance for each sample (Pedregosa et al., 2011). As the best value of this metric is 1, we want to maximise it.

Finally, in DB metric lower values indicates better clustering, so we want to minimise it. The formulation of DB score can be found in Davies & Bouldin (1979).

Clustering results

Table 1 and Table 2 show the results of the clustering performance evaluation metrics of the clusters obtained with the k-means and the hierarchical clustering algorithms, respectively. The clusters obtained using the three ICs outperform the rest of attributes combinations using two of the metrics. As we can see, when the original dataset is used, the values obtained using the PBM metric are much higher than when the dataset without the outlier is used. This is due to the influence of the outlier in that metric. Regarding the combination of original attributes, better clusters are obtained using 5 attributes. Nevertheless, better results are obtained using the 3 ICs, as metrics SIL and DB show. Based on that, we selected the clusters obtained using the original dataset with five clusters (k=5) and 3 ICs. Figure 4 shows the cluster assignments obtained and used for the energy forecast models. In that graph, the cluster outlier corresponding to the last week of June is assigned to an independent cluster (labelled as “50”). We could think that the cluster we labelled as “10” could correspond to winter conditions, as the first and next to last weeks of the year was assigned to it. The cluster “40” can be interpreted as summer as the middle part of the year was assigned to that cluster. Then, the other two clusters, labelled as “20” and “30” could be autumn and spring and different weeks during the year were assigned to them. Each of the identified clusters represent different environmental conditions in which the building could be operating under different settings and modes. New environmental data can be classified in the nearest cluster by calculating the distance to the cluster centroid. Then, the energy demand model specifically built for that cluster will be used to predict the energy demand.

Table 1. Clustering performance evaluation of k-means clusters. For metrics SIL and PBM higher is better and for metric DB lower is better.

K-means	*Original dataset*		*Dataset without outlier*		*Metric*
	k=5	**k=4**	**k=4**	**k=3**	
3 attributes	0.10	0.12	0.11	0.12	*SIL*
	1.73	1.65	2.06	2.07	*DB*
	4.82	6.49	1.68	2.45	*PBM*
4 attributes	0.18	0.19	0.18	0.17	*SIL*
	1.21	1.40	1.46	1.68	*DB*
	5.45	2.68	2.37	1.68	*PBM*
5 attributes	0.21	0.22	0.21	0.21	*SIL*
	1.15	1.32	1.35	1.46	*DB*
	5.96	3.31	2.49	2.18	*PBM*
3 ICs	**0.30**	**0.31**	**0.33**	**0.31**	*SIL*
	0.85	**0.84**	**0.98**	**1.12**	*DB*
	3.83	4.34	1.66	1.53	*PBM*
4 ICs	0.22	0.22	0.25	0.23	*SIL*
	1.08	1.10	1.23	1.47	*DB*
	2.61	3.35	1.08	0.88	*PBM*

Table 2. Clustering performance evaluation of clusters obtained with hierarchical clustering. For metrics SIL and PBM higher is better and for metric DB lower is better.

Hierarchical clustering	*Original dataset*		*Dataset without outlier*		*Metric*
	k=5	**k=4**	**k=4**	**k=3**	
3 attributes	0.10	0.13	0.1	0.14	*SIL*
	1.68	1.54	1.99	1.92	*DB*
	4.68	6.89	1.61	2.70	*PBM*
4 attributes	0.13	0.14	0.14	0.16	*SIL*
	1.31	1.53	1.54	1.73	*DB*
	5.12	2.02	1.80	1.85	*PBM*
5 attributes	0.17	0.18	0.18	0.20	*SIL*
	1.17	1.34	1.37	1.52	*DB*
	5.73	2.25	2.01	2.16	*PBM*
3 ICs	**0.31**	**0.28**	**0.30**	**0.29**	*SIL*
	0.83	**0.92**	**1.05**	**1.26**	*DB*
	4.91	4.24	1.67	1.28	*PBM*
4 ICs	0.21	0.21	0.21	0.19	*SIL*
	1.10	1.34	1.33	1.61	*DB*
	2.95	0.86	1.10	0.86	*PBM*

Energy forecasting models using LSTM

The second phase of the methodology is to develop two energy forecasting models and compare their performance. One model ("cluster model") was specifically constructed for energy forecasting in one of the clusters identified. The other model ("general model") is constructed to forecast the energy consumption during the whole year, without considering the operational state of the system.

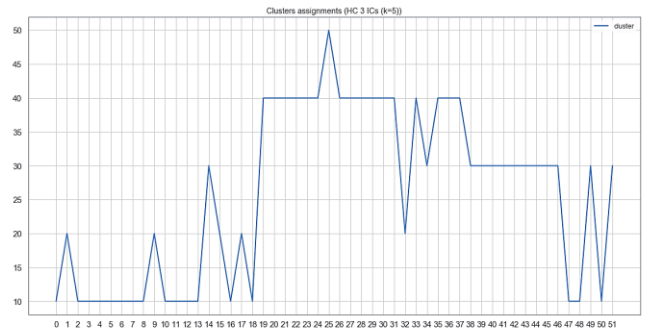

Figure 4. Weekly data cluster assignations using hierarchical clustering, 3 ICs and five clusters (k=5).

Four weeks (weeks 21 to 24 of Figure 4) assigned to the same cluster were chosen to construct the specific cluster model. The first three weeks were used for training (Figure 5) and the last one for testing (Figure 6). The same test dataset was reserved and used for the general mode. The energy demand of the previous timestep, the day of the year and the hour were the inputs for the models.

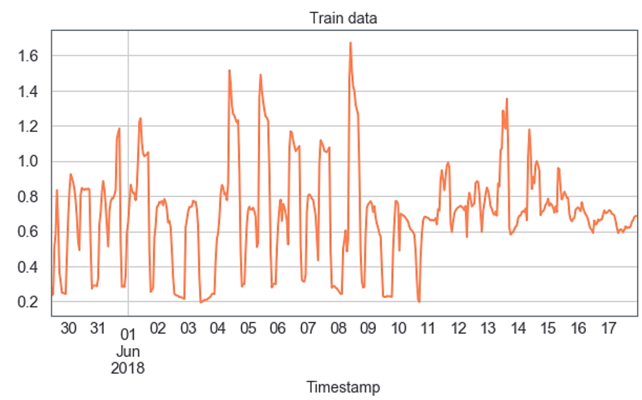

Figure 5. Train dataset for the cluster model.

We validated the model using 30% of the training data. The data was cleaned and standardized to have zero mean and a standard deviation of one using the z-score formula. For this standardization we calculated the mean and the standard deviation using only the training data, simulating a real scenario where we know nothing about the test data. Then, we standardize all the data based on those calculated metrics.

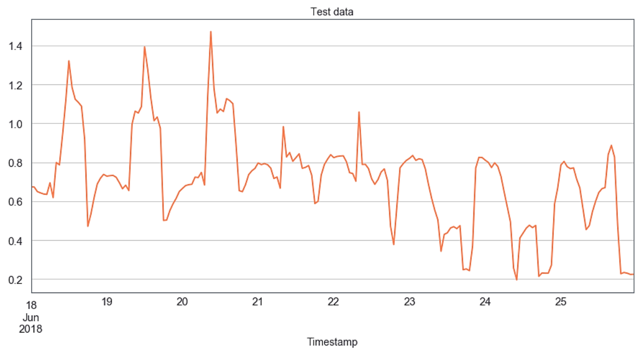

Figure 6. Test dataset for both the cluster and the general model.

We selected the LSTM algorithm to construct the models for building level energy forecasting. For the implementation of the LSTM models we used the high-level neural networks API Keras (Chollet & others, 2015) running on the top of TensorFlow (Martín Abadi et al., 2015). We used hourly data of the total thermal energy demand for the heating coils of the AHUs. The stateless LSTM network was fed with five timesteps back in order to predict the next. For training the model, we used ADAM (Kingma & Ba, 2017) algorithm because of its fast convergence. The models were composed of two layers with 100 and 50 hidden neurons, respectively. To determine the right number of epochs and batch sizes, we performed several tests with the models. The best cluster model was obtained with 40 epochs and a batch size of 16, and the best general model was obtained with 30 epochs and 20 as batch size. To decrease the overfitting of the model to the training data and increase it accuracy on the test dataset, we used Dropout as regularization methodology (Srivastava et al., 2014).

Results and discussion

We adopted Root mean squared error (RMSE) as performance metric to assess the models obtained.

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=0}^{N}(\hat{y}_i[t] - y_i[t])^2}$$

We compared the forecasting accuracy of the cluster model with the general model. Figure 8 illustrates the performance of the cluster model in the test dataset. In this model, we obtained a RMSE of 0.125 in the test data, while using the general model (Figure 7), we obtained a RMSE of 0.205. As shown in those two figures, the lowest performance of the model is achieved on the weekends (June 23rd and 24th, 2018) especially in the general model due to the high variability of the data in them. We can see in this test, the usefulness of the clustering previously done to identify the different operational modes of the building. From these results, we can say that our initial guess was right and using clusters to identify operational models and creating different models for each cluster is better than having one general model.

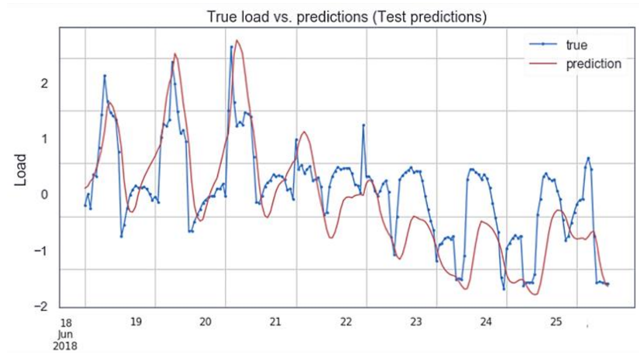

Figure 7. Prediction results and real measurements for the test dataset using the general model.

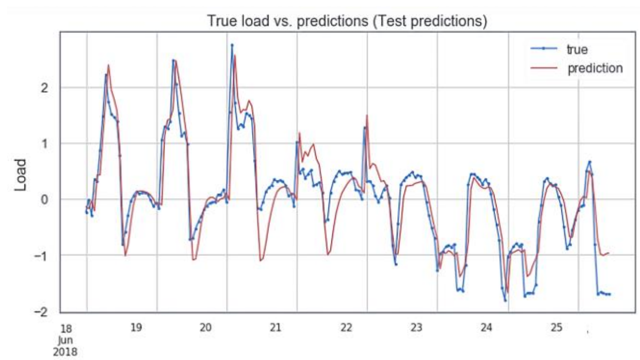

Figure 8. Prediction results and real measurements for the test dataset using the cluster model.

Summary and conclusions

In this paper, we presented an innovative methodology to address the issue of building level energy forecasting. We analysed the weather data attributes, not only in their original values but also using ICs. By using that environmental weather data, we characterised the different operational modes, in which the building operates in the form of five clusters. Two types of clustering methods were applied, k-means and hierarchical clustering. Finally, five clusters were clearly identified, one corresponding to an unusual hot and non-humid week at the end of June and the remaining corresponding to the four different weather seasons.

A baseline model (general model) for energy forecasting was constructed using the energy consumption data for the whole year and it was compared with a specialised model (cluster models) constructed for a specific building operation mode. Then we have shown that the cluster model outperforms the general model.

As future work, we will define and assess the methodology to perform predictive maintenance based on those cluster models. We will also continue to investigate and compare the performance of other types of deep learning architectures in the energy consumption forecasting. A focus for the future work will be also on getting accurate long-term energy consumption forecasting. Finally, to be able to better generalise the methodology proposed in this paper, more datasets and timeframes will be tested using the proposed approach.

Acknowledgment

The data used in the case study of this paper was provided by the IRUSE team and the College of Informatics and Engineering of NUI Galway. This research work was funded by the NUI Galway CoEI Postgraduate Scholarship 2018 and the European Commission under grant agreement number 820805 through SPHERE project.

Nomenclature

AdaBoost	Adaptative Boosting
ADAM	A Method of Stochastic Optimization
AHU	Air Handling Unit
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average
BP	Back Propagation
DB	Davies-Bouldin
HVAC	Heating, Ventilation and Air Conditioning
ICA	Independent Component Analysis
ICs	Independent Components
LSTM	Long-short Term Memory
NUIG	National University of Ireland, Galway
PBM	Pakhira, Bandyopadhyay and Maulik
RBM	Restricted Boltzmann Machine
RH	Relative Humidity
RMSE	Root mean squared error
RNN	Recurrent Neural Network
SIL	Silhouette
SVM	Support Vector Machine
WCSS	Within Clusters Summed Squares

References

Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2018). A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. *Energy and Buildings*, *165*, 301–320. https://doi.org/10.1016/j.enbuild.2018.01.017

Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., & Jenssen, R. (2017). An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting. *ArXiv:1705.04378 [Cs]*. https://doi.org/10.1007/978-3-319-70338-1

Chollet, F. & others. (2015). *Keras*. https://keras.io

Costa, A., Keane, M. M., Torrens, J. I., & Corry, E. (2013). Building operation and energy performance: Monitoring, analysis and optimisation toolkit. *Applied Energy*, *101*, 310–316. https://doi.org/10.1016/j.apenergy.2011.10.037

Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-1*(2), 224–227. https://doi.org/10.1109/TPAMI.1979.4766909

Gasparin, A., Lukovic, S., & Alippi, C. (2019). *Deep Learning for Time Series Forecasting: The Electric Load Case*. 19.

Graves, A., Mohamed, A., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 6645–6649. https://doi.org/10.1109/ICASSP.2013.6638947

Habernal, I., & Matoušek, V. (Eds.). (2013). *Text, Speech, and Dialogue: 16th International Conference, TSD 2013, Pilsen, Czech Republic, September 1-5, 2013. Proceedings* (Vol. 8082). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40585-3

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hyvarinen, A., & Oja, E. (1997). *A Fast Fixed-Point Algorithm for Independent Component Analysis*. 10.

International Energy Agency. (2018a). *CO2 Emissions from Fuel Combustion 2018 Highlights*. 166.

International Energy Agency. (2018b). Energy Efficiency Indicators 2018: Highlights. *Energy Efficiency*, 191.

Kingma, D. P., & Ba, J. (2017). Adam: A Method for Stochastic Optimization. *ArXiv:1412.6980 [Cs]*. http://arxiv.org/abs/1412.6980

Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using Deep Neural Networks. *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, 7046–7051. https://doi.org/10.1109/IECON.2016.7793413

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, … Xiaoqiang Zheng. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. https://www.tensorflow.org/

Medojevic, M., Díaz Villar, P., Cosic, I., Rikalovic, A., Sremcev, N., & Lazarevic, M. (2018). *Energy management in industry 4.0 ecosystem: A review on possibilities and concerns*. *29*, 0674–0680. Scopus. https://doi.org/10.2507/29th.daaam.proceedings.097

Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2016). Deep learning for estimating building energy consumption. *Sustainable Energy Grids & Networks*, *6*, 91–99. https://doi.org/10.1016/j.segan.2016.02.005

Naug, A., & Biswas, G. (2018). Data Driven Methods for Energy Reduction in Large Buildings. *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*, 131–138. https://doi.org/10.1109/SMARTCOMP.2018.00083

Naug, A., Quiñones-Grueiro, M., & Biswas, G. (2020). A Relearning Approach to Reinforcement Learning for Control of Smart Buildings. *ArXiv:2008.01879 [Cs, Eess]*. http://arxiv.org/abs/2008.01879

Ozawa, S., & Kotani, M. (2000). *A Study of Feature Extraction and Selection Using Independent Component Analysis*. 6.

Pakhira, M. K., Bandyopadhyay, S., & Maulik, U. (2004). Validity index for crisp and fuzzy clusters. *Pattern Recognition*, *37*(3), 487–501. https://doi.org/10.1016/j.patcog.2003.06.005

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2825–2830.

Rokach, L., & Maimon, O. (2005). Clustering Methods. In O. Maimon & L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook* (pp. 321–352). Springer-Verlag. https://doi.org/10.1007/0-387-25465-X_15

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Sterling, R., Struß, P., Febres, J., Sabir, U., & Keane, M. (2014). *From Modelica Models to Fault Diagnosis in Air Handling Units*. 447–454. https://doi.org/10.3384/ecp14096447

Warburton, P., Butcher, K. J., & Chartered Institution of Building Services Engineers (Eds.). (2009). *Building control systems* (2. ed). CIBSE.

Zhou, C., Fang, Z., Xu, X., Zhang, X., Ding, Y., Jiang, X., & ji, Y. (2020). Using long short-term memory networks to predict energy consumption of air-conditioning systems. *Sustainable Cities and Society*, *55*, 102000. https://doi.org/10.1016/j.scs.2019.102000