Forecasting building energy performance using machine learning methods

Mahnameh Taheri, Colin Parry, Agnieszka Hermanowicz, Alan Wegienka
arbnco Ltd., Glasgow, United Kingdom

Abstract

Rapid and reliable energy performance predictions using building performance simulations (BPS) has been one of the main concerns of the building science community. Machine learning techniques use nonlinear regression models that are trained to quickly approximate the building performance. This contribution explores the utility of dilated convolutional neural networks (dCNNs) for forecasting time-series of energy data. Energy and weather data are used for model training. The performance of dCNNs is examined in terms of efficiency and accuracy for forecasting time series of data. This paper explores whether, for certain applications, similar techniques could be used as an alternative to BPS.

Introduction

Building performance simulation offers a numerical approximation of the thermo-physical processes in and around a building. For accurate and reliable simulation results, accurate and detailed inputs are required, pertaining to the building, its boundary conditions, operational patterns, etc. Moreover, the respective calculations involve an exhaustive computation over fine discretisation in space and time. Therefore, targeting reliable and rapid predictions remain one of the main objectives in development of building performance simulation tools (Luo et al. 2017). Some use cases of building performance evaluations are, however, based on low specialist involvement and minimum building information.

To provide usable insights from high-level data, methodologies have been proposed based on standardised reference building models, or regression models trained on monitored data. In fact, working directly with actual energy consumption data from the targeted building and regression models have the potential to generate significant insight. The most popular regression analyses are based on linear models and monthly or yearly heating and cooling degree days (Fels 1986). Weather normalization is in fact the most commonly used form of predicting energy consumption. However, temperature is only one of the weather components and there are other parameters, such as wind speed, humidity, and solar radiation, which also influence the building energy performance (Dong et al. 2005). This, together with other uncertainties involved in the degree day calculations limits the scope and accuracy of this method. Improving the building energy performance predictions and addressing the shortcomings of the conventional approaches encouraged the development of new techniques. More sophisticated black-box models have been developed to imitate the causal relationships between the energy performance of the building and a set of independent variables using Machine Learning (ML) methods. Despite the calculations in BPS, the causal relationships here are not broken down to physics-based rules (Luo et al. 2017).

Forecasting time series using ML

Forecasting time series of data based on past observations is clearly of interest in many fields. However, the non-linear trends, heavy tails and noise make the accurate prediction of the temporal relationships difficult (Borovykh et al. 2018). Traditionally, autoregressive techniques such as autoregressive moving-average (ARMA) have been used for time series forecasting (Hamilton 1994). However, these techniques do poorly when the relationship between independent and dependent is non-linear. The current standard for non-linear time series modelling is recurrent neural networks (RNNs) (Williams et al. 1986), specifically the long-short term memory (LSTM) unit (Hochreiter et al. 1997). These learn from the data sequentially making them very powerful but also very slow to train.

Convolutional neural network (CNN) is a type of deep neural network which gained initial traction in image classification problems (Aussem et al. 1997, Krizhevsky et al. 2012) but has significantly branched out into other problem areas, such as time series forecasting. A CNN works by applying a series of sliding transformations to the data to reduce its complexity and learn from the underlying trend. The weights applied during the transformations become the parameters learned during model fitting.

A dilated CNN (dCNN) takes advantage of the convolution by applying a series of layers whose neurons have two neurons from a previous layer as the input. Therefore, the receptive field grows exponentially as more layers are added (Figure 1). By increasing the number of dilation layers the model can learn from larger amounts of historical data without exponentially increasing the number of neurons, so

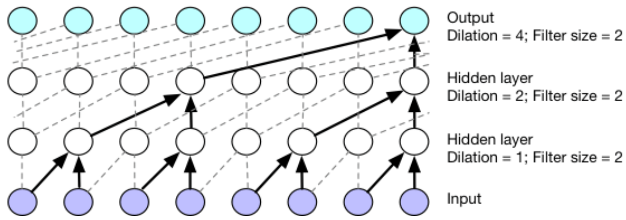

Figure 1: A dCNN architecture with two hidden layers resulting in a receptive field of eight previous points.

training time is dramatically faster than using a RNN based architecture.

With this background the current contribution explores the utility of dCNN, in terms of efficiency and accuracy, for forecasting time series of building energy data.

Methods

Database

In a previous study the authors used an online library of data from non-residential buildings, i.e., the genome project, published by Miller and Meggers (2017). The library includes one year of hourly, whole-building electrical meter data for 507 non-residential buildings, together with building size and the weather file (Miller and Meggers 2017). The required time and effort for data collection, processing and replicability can be reduced by using the standard test database (Tenopir et al. 2011, Taheri and Mahdavi 2018). Systematic ontologies and standardized data format have been developed for the benefit of assisting universal data sharing (for example, see, Mahdavi and Taheri 2017). However, unfortunately the mentioned online library could not be used in this work, as the minimum amount of data required to train the dCNN model is twelve months of hourly or better data. Thus, this data set was insufficient for validation purposes. Instead, the dataset for training and evaluation used in this study is a series of anonymised building energy time series, from a range of different building types. The building types include office, hotel, leisure centre, farm, manufacturing, and food producing facilities, located in the UK and US (see, Table 1). For the purpose of this contribution presenting the details of the data is limited to what is included in Table 1. Each energy time series has multiple years of data. To maximise the amount of data available, the data from each building was split into multiple twelve months windows with 3 months for validation. Each window moved forward 3 months (so that 9 months of training data overlapped, but the validation data did not, see Figure 2). After removal of time series windows with too much missing data, gave a total of 243 time series.

Table 1: The start and end dates of each building energy time series, along with total duration and sampling rate.

ID	Building Usage	Start data	End data	Duration (Days)	Sampling Rate (min)
1	Office A	02-01-01	01-01-04	1093	15
2	Office B	02-01-01	01-01-04	1093	15
3	Office C	10-04-16	30-04-19	1114	30
4	Office D	30-04-15	30-04-19	1460	30
5	Office E	26-02-16	30-04-19	1159	30
6	Hotel A	01-01-15	31-03-19	1550	30
7	Hotel B	01-01-15	30-04-19	1580	30
8	Hotel C	01-01-15	30-04-19	1580	30
9	Leisure A	01-01-15	30-04-19	1580	30
10	Leisure B	01-01-15	30-04-19	1580	30
11	Leisure C	01-01-15	30-04-19	1580	30
12	Leisure D	01-01-15	30-04-19	1580	30
13	Leisure E	01-01-15	30-04-19	1580	30
14	Leisure F	01-01-15	30-04-19	1580	30
15	Leisure G	01-01-15	30-04-19	1580	30
16	Leisure H	01-01-15	30-04-19	1580	30
17	Leisure I	01-01-15	30-04-19	1580	30
18	Food Producing A	01-01-15	30-04-19	1579	30
19	Food Producing B	01-01-15	30-04-19	1580	30
20	Manufacturing A	01-01-15	30-04-19	1580	30
21	Manufacturing B	01-01-15	30-04-19	1580	30
22	Manufacturing C	01-01-15	30-04-19	1580	30
23	Farm	01-01-15	30-04-19	1580	30

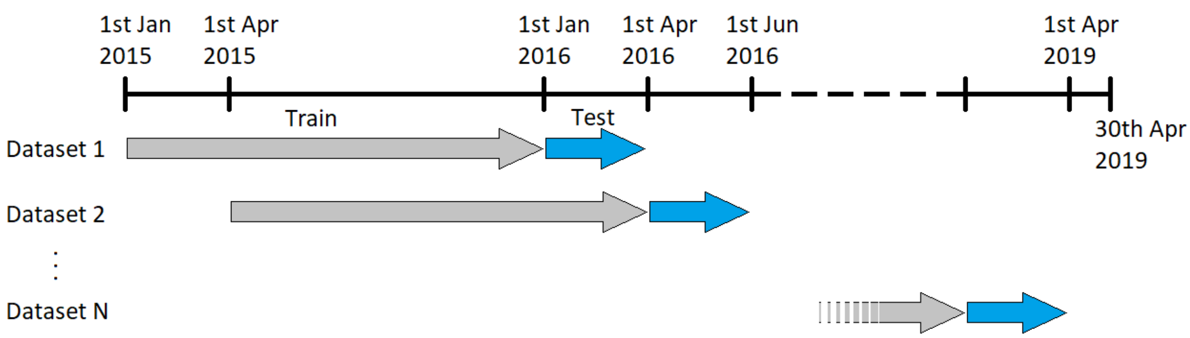

Figure 2: Splitting a long time series into 12 months of training and 3 months of test data multiple times. ID 9 from Table 1 was used as an example.

Independent variables

As mentioned above, the so-called black-box model is expected to replicate the causal relationships between weather and non-weather parameters, and the energy consumption, entirely by mathematical relationships learnt from data. Below is the list of independent parameters used in this study:

- Weather-dependant variables, i.e., outdoor temperature and relative humidity: Weather data has been downloaded from Dark Sky, for the location of the buildings and the respective time periods (Dark Sky 2012).
- Non-weather-dependant variable, i.e., operational patterns captured from data: The operational patterns are an attempt to approximate occupancy of the building, which is not known. In order to do this the data is first grouped into periods of similar operation by applying a Gaussian Mixture Model (GMM) to the daily base load (15th percentile) and obtaining up to four periods (see Figure 3). Any periods with an insufficient number of points are removed and these points assigned to the next most similar period. Then, a representative weekly profile for each period is created by calculating the mean load for each time interval of the week. Since different periods can have different magnitudes of energy, these values are normalised between 0 and 1 using a threshold value and the 85th percentile of energy in the period respectively (Luo et al. 2017). This is illustrated in Figure 4 . The threshold value E_{thresh} is calculated using:

$$E_{thresh} = E_{p5} + \frac{1}{4}\left(E_{p95} - E_{p5}\right) \qquad (1)$$

Where:

E_{p5} and E_{p95} are the 5th and 95th percentile of energy in the given cluster.

This representative weekly profile, calculated per period of similar operation, is a useful model input.

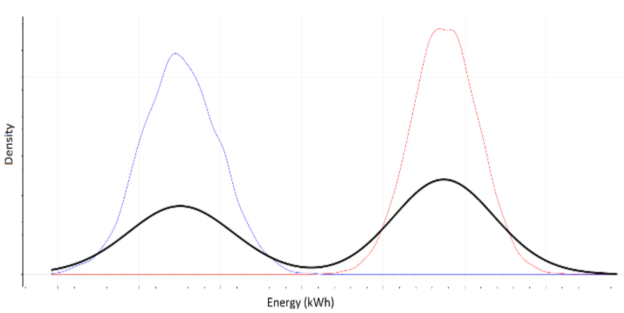

Figure 3: The distribution of base loads for a building (black) with two fitted Gaussian Mixture Models (blue and red) imposed on top. These models can then be used to classify days into each period based on their probability of association.

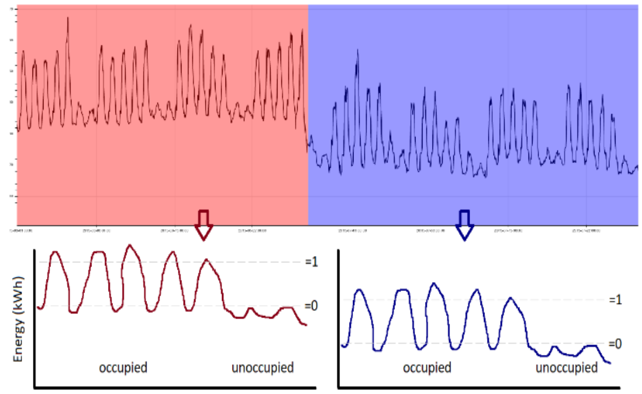

Figure 4 : Once the days have been grouped together using base load, the mean load at each time step is calculated and normalised between 0 and 1.

Model Architecture

The dilated CNN architecture used in this paper is shown in Figure 5. The input is a series of sequences which the neural network learns from to predict a series of sequences at the output block. The main building block for this model architecture is known as a convolution. Put simply a convolution is the method of taking a filter, which is much smaller than the original dataset, and multiplying its weights with the original values and summing them to get a single value. The process of sliding, or convolving, the filter across the data gives it its name. This step allows the model to simplify the input data and learn the overall trend.

The main convolutions take place between the initial and final dilated 1D convolution blocks, and the optimal receptive field is found to be around three weeks of historical data. A final skip connections block (Long et al. 2015) allows the coarse output of the final convolution to be combined with the much finer output of the initial convolution. A final 1D convolution allows this combined output to be further combined and simplified.

Neural network architectures can very quickly become very complex so that the model learns the statistical noise in the dataset and then does not generalise well to new data. One way to reduce this overfitting is to randomly hide neurons during the training process so that each weight update does not include all neurons – this is known as dropout (Srivastava et al. 2014). This forces neurons to generalise better since their weights are updated irregularly. During the prediction stage all neurons are utilised, the dropout only applies during training. Finally, the model architecture is flattened and reduced down to the final output shape. Predicting a sequence of around one day of data is found to give good results.

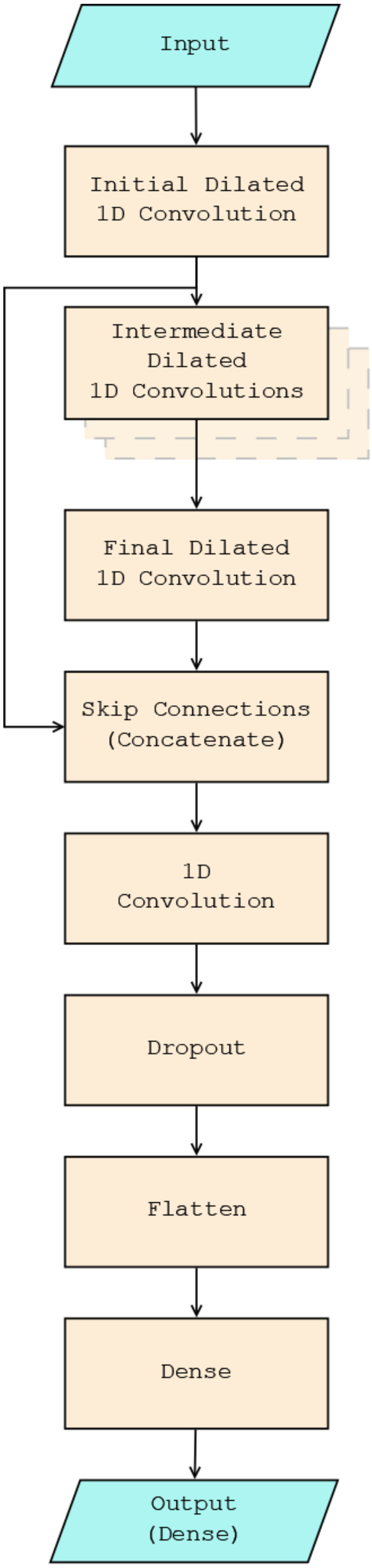

Figure 5: The building blocks which make up the dilated CNN architecture used in this paper.

Uncertainty Calculations

Two principal uncertainty indices, proposed by ASHRAE guidelines, have been used for the measure of the forecast uncertainties in this study. These include, Normalized Mean Bias Error (NMBE), and Coefficient of Variation of the Root Mean Square Error (CV(RMSE)) (ASHRAE 2014, Ruiz and Bandera 2017).

$$CV(RMSE) = 100 \frac{\sqrt{\frac{(y_i - \hat{y}_i)^2}{(n-p)}}}{\bar{y}} \quad (2)$$

$$NMBE = 100 \frac{\sum(y_i - \hat{y}_i)}{(n-p) \times \bar{y}} \quad (3)$$

Where:

y_i= Measured data point

$\hat{y}_i$= Simulated data point

$\bar{y}$ = Mean of measured data points

n = number of data points (e.g., 8760 for hourly)

p = number of parametric outputs.

CV(RMSE) "aggregates the runtime individual time step errors into a single dimensionless number" (Taheri et al. 2013). MBE, the average of the errors of a sample space, is a good indicator of the overall behaviour of the simulated data regarding the regression line of the sample. Positive and negative values here indicate the under- or over-predictions. NMBE is used to make the results of MBE comparable.

Results

The box plots in Figure 6 demonstrate the models evaluation statistics for the 243 datasets.

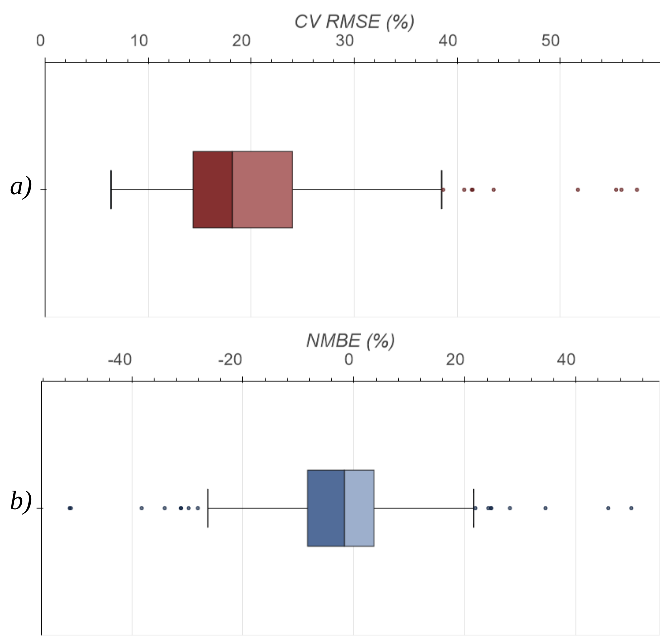

Figure 6: Distribution of model evaluation statistics for 243 tested datasets, a) CV(RMSD), and b) NMBE middle.

According to ASHRAE Guide 14, for building energy simulation the computer model requires a Normalized Mean Bias Error (NMBE) and a Coefficient of Variation of the Root Mean Square Error CV(RMSE) of below 5% and 15%, respectively, for monthly data. If hourly data are used, these requirements shall be 10% and 30%, respectively. According to Table 1 the resolution of the data used in this study is half-hourly or better. Considering similar level of errors as suggested by ASHRAE, the average calculated errors presented in Figure 6a and b are considered acceptable (ASHRAE 2014). Figure 7 and Figure 8 demonstrate the time series of actual energy data vs. the predicted energy data in training and test periods. The yellow line is the original data, the grey is the learned training data and blue is the predicted future data. Note that, looking at the charts there is a small gap in the predicted data both for the training and test predictions. This gap is caused by the size of the receptive field required to make the first prediction. As the sequences used to predict partially overlap with each other, the test set was only evaluated on test data which did not use any training data to predict, i.e., one receptive field from the first timestamp. The graphs in Figure 7 demonstrate the power of the model in capturing the daily and seasonal consumption pattern in the energy consumption. However, examples of Figure 8 are failing in prediction.

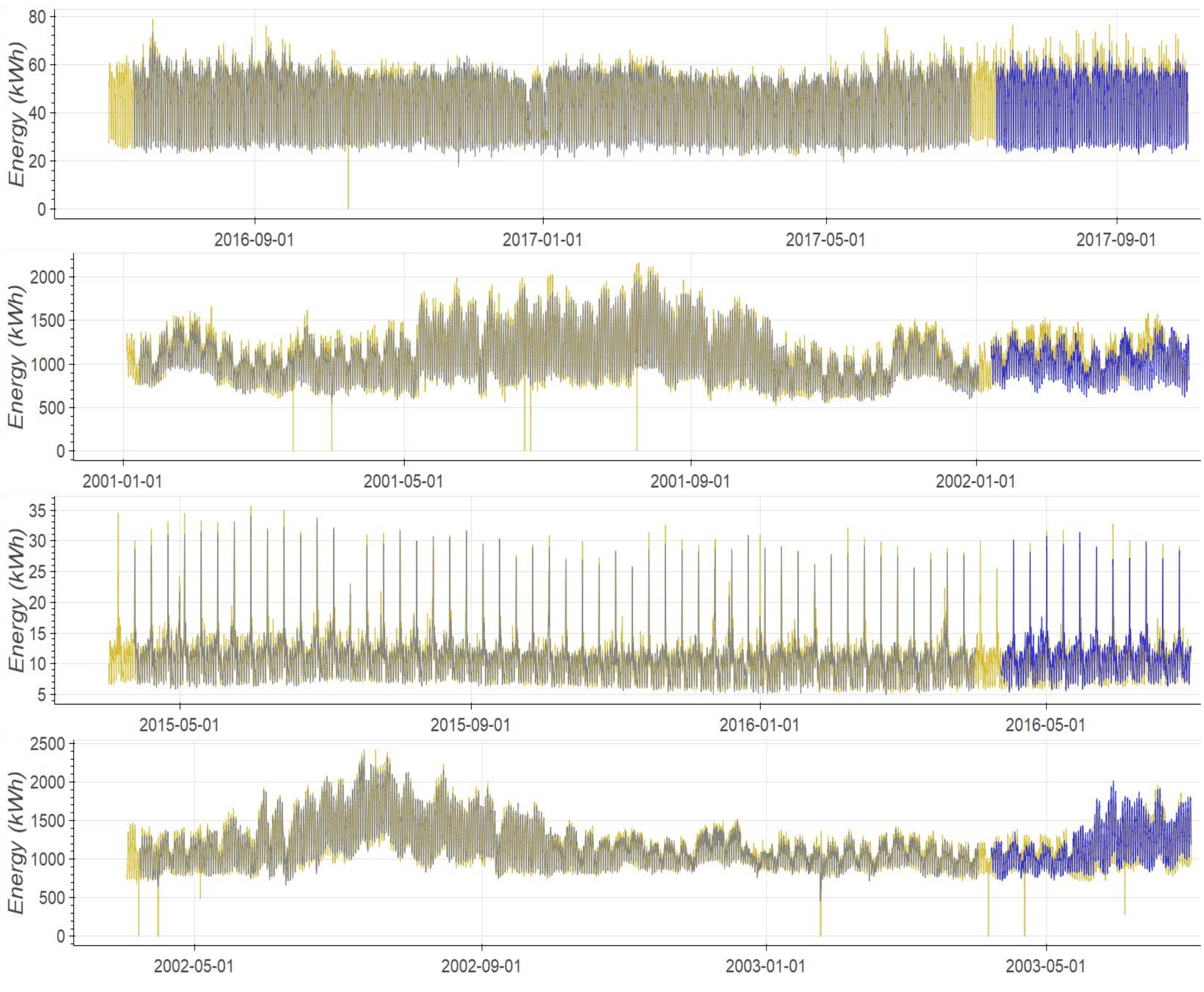

Figure 7 Time series of actual energy data vs. the predicted energy data of training and test periods – four well performing examples. Original data, in yellow, learned training data, in grey, predicted future data, in blue.

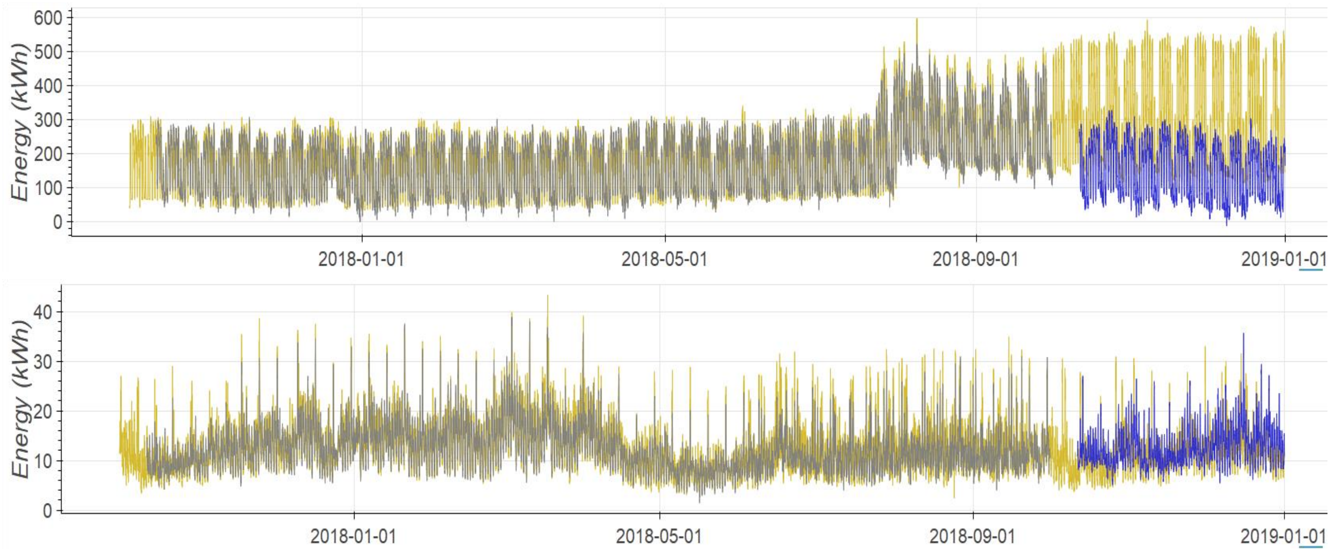

Figure 8: Time series of actual energy data vs. the predicted energy data of training and test periods - two bad performing examples. Colours the same as Figure 7.

Further investigation of the results suggests that the energy consumption pattern in some buildings does not follow any particular pattern. Without access to more information about the building, its occupants, and operation strategies, the reason behind is unclear. This, in part, can be explained by the diversity of commercial buildings' inhabitants and their influence on buildings' energy use over the time. Change of buildings' use case and inhabitants' different sensitivities and preferences regarding comfort condition could be behind buildings disparate response to the same environmental condition. Thus, the model cannot learn the pattern from the data. Moreover, significant change in the energy consumption as compared to the similar period from previous year is another reason preventing an acceptable prediction.

Figure 9 shows a plot of CV(RMSE) versus NMBE for the test data. Having a low CV(RMSE) with around zero bias, with a few exceptions, shows the prediction magnitude is appropriate and follows the trend of the real data. It can be seen that there is a relationship between the two, which is due in part to the worst performing time series having a test data set with a change in usage not present in the training data set.

The box plot in Figure 10 shows the distribution of the training times for the models. The models were trained using the Keras API through Tensorflow 2.1.0 on a machine with a NVIDIA GeForce GTX1080Ti GPU. It can be seen that the majority of models trained in under 21 seconds, which was considered fast enough for online training.

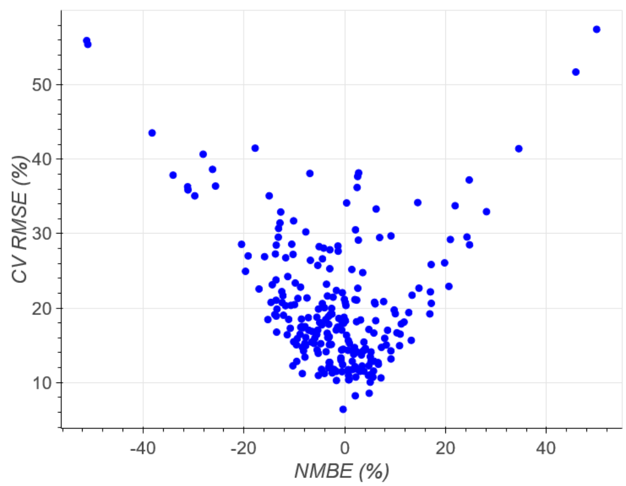

Figure 9: A plot of CV(RMSE) vs NMBE for the test data

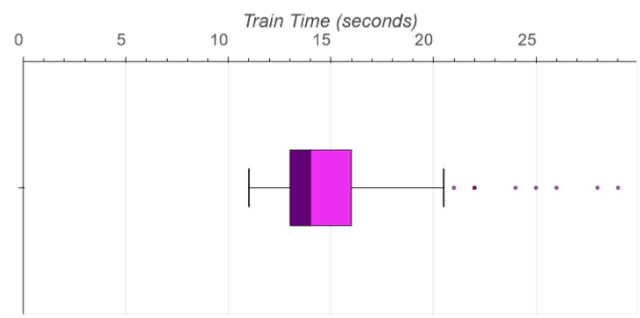

Figure 10: Distribution of model train time for 243 tested datasets.

Discussion

Below is the list of some of the relevant discussions and future considerations:

- Normally when training a machine learning model, the weights are learned offline from a known dataset and applied to future predictions. However, we expect in cases similar to the problem presented here, the response of each building to the independent variables differs widely. This depends on various reasons, including for instance, the physics of building, the building usage type, etc. As a result, in this study the model is trained online on a per building basis. Therefore, a crucial parameter during the model evaluation phase was the duration of the training time. The fact that the dCNN learns from an exponentially increasing amount of historical data, without the training time being affected exponentially, makes it a perfect candidate for the present use case.
- The hyperparameters were evaluated across a range of different building types, by not only evaluating the model's accuracy and bias, but also the training time. A few complex models with too large dilations were removed for this reason.
- The minimum amount of data required to train the dCNN model is 12 months of hourly or better data. Since the aim in this study is that the users of this method only need twelve months of data, when the models are trained online there is not enough surplus data to create a validation set. Therefore, the early stopping criteria only takes account of the training loss and has to be carefully optimised to prevent overfitting.
- At the initial stage of the study an attempt was made to consider solar irradiance as the fourth independent variable for the model. For this purpose, models such as Zhang-Huang model were required to convert the cloud cover available in the weather data into the solar irradiance (Zhang and Huang 2002). However, reliability of the cloud cover data for all locations could have been under question. Therefore, the authors did not include the solar irradiance as an independent parameter in order not to add more uncertainties to the problem. In the future iteration of the model the application of simulated solar irradiance will be used to test its impact on improving the results.

Conclusion

To meet the new ambition of net-zero carbon emissions the governments approach to retrofitting existing buildings requires a rapid scale up. Considering that 85%-95% of the buildings in 2050 have already been built today emphasises the importance of fast and reliable performance evaluations of the buildings in their operation phase. Running building energy performance simulation involves accurate and comprehensive input data, about the building and its surroundings. In contrast, Machine Learning (ML) techniques require minimal building information for energy modelling. Here, apart from the laws of physics the causal relationships between building's energy performance and its external conditions are modelled by sophisticated mathematical relationships. These relationships and interactions are purely learned from the data itself. This contribution explored the utility of dilated convolutional neural networks (dCNNs), for forecasting time series of energy data. Towards this end, actual metered energy data of a set of available buildings as well as the weather conditions has been used for model training. The method here requires availability of twelve months of data. Since the model is trained based on the correlation between energy consumption, outdoor weather condition, and operational patterns, a full year of data gives a better picture of considerable changes in the operational pattern during different time periods in a year. We examined the performance of dCNNs against the holdout dataset in terms of efficiency and accuracy for forecasting time series of energy data. Note that, the number of tested buildings (243 datasets, generated based on the data from 23 buildings) and the rather limited list of building types pose limits on a very general validity of the results. However, the method presented in this paper is based on online model training on a per building basis. This approach makes it applicable for buildings from different usage types and with widely different response patterns to the aforementioned independent variables. According to the findings the presented approach provides reasonable accuracy required for wide range of building performance assessment applications.

Acknowledgement

The authors would like to acknowledge the contributions of arbnco's Energy Product team members in the discussions related to the research presented in this paper. The authors would also like to thank Maureen Eisbrenner and Simon West for their support.

References

American Society of Heating, Refrigerating and Air-Conditioning Engineers, ASHRAE . (2014). Guideline 14-2014, Measurement of Energy and Demand Savings; Technical Report; American Society of Heating, Ventilating, and Air Conditioning Engineers: Atlanta, GA, USA, 2014.

Aussem, A. and F. Murtagh (1977) Combining neural network forecasts on wavelet-transformed time series. *Connection Science 9*, 113–122.

Borovykh, A., Bothe, S., Oosterlee, C.W. (2018) Dilated Convolutional Neural Networks for Time Series Forecasting. Journal of Computational Finance. DOI: 10.21314/JCF.2019.358.

Dark Sky. 2012. The Dark Sky Company, LLC. Available on https://darksky.net/. Last accessed 18.02.2020.

Dong, B., Lee, S. E., Haji Sapar, M. (2005). A Holistic Utility Bill Analysis Method for Baselining Whole Commercial Building Energy Consumption in Singapore. *Energy and Buildings 37*, 167–174.

Hamilton, J.D. (1994) Time series analysis, vol. 2, Princeton university press Princeton, 1994.

Hochreiter, S. and J. Schmidhuber (1997) Long Short-Term Memory. *Neural computation 9*, 1735–1780.

Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25*, 1097–1105.

Long, J., Shelhamer, E., Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. Available online: https://arxiv.org/pdf/1411.4038.pdf. Last accessed: 12.02.2020.

Luo, X., Hong, T., Chen, Y., Piette, M. (2017). Electric Load Shape Benchmarking for Small- and Medium-Sized Commercial Buildings. Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.

Mahdavi, A. and M. Taheri (2017). An Ontology for Building Monitoring. *Journal of Building Performance Simulation 10* (5-6). DOI: 10.1080/19401493.2016.1243730.

Miller, C. and F. Meggers (2017). The Building Data Genome Project: An Open, Public Data Set from Non-Residential Building Electrical Meters. *Energy Procedia 122*, September 2017, 439-444.

Ruiz, G.R. and C.F. Bandera (2017). Validation of Calibrated Energy Models: Common Errors. *Energies 2017*, 10, 1587. Doi:10.3390/en10101587.

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

Taheri, M. and A. Mahdavi (2018). Structured Representation of Monitored Occupancy Data. *Bauphysik 40* (6).

Taheri, M., Tahmasebi, F., Mahdavi, A. (2013). A case study of optimization-aided thermal building performance simulation calibration. Proceedings of the 13th International Conference of the International Building Performance Simulation Association, Chambéry, France.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., et al. (2011) Data Sharing by Scientists: Practices and Perceptions. PLoS ONE 6(6): e21101. https://doi.org/10.1371/journal.pone.0021101.

Williams, Ronald J.; Hinton, Geoffrey E.; Rumelhart, David E. (1986) Learning representations by back-propagating errors. *Nature 323* (6088) 533–536.

Zhang, Q.Y. and Y.J. Huang (2002). Development of Typical Year Weather Files for Chinese Locations, LBNL-51436, ASHRAE Transactions, Vol. 108, Part 2.