Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 161 (2019) 242–250

Energy Procedia

www.elsevier.com/locate/procedia

2nd International Conference on Sustainable Energy and Resource Use in Food Chains, ICSEF 2018, 17-19 October 2019, Paphos, Cyprus

A data-driven approach for electricity load profile prediction of new supermarkets

Ramon Granell[a, b]*, Colin J. Axon[a], Maria Kolokotroni[a] David C.H. Wallom[b]

[a]*Brunel University London, Institute of Energy Futures, Uxbridge UB8 3PH, United Kingdom*
[b]*Oxford e-Research Centre, Department of Engineering Science, University of Oxford, 7 Keble Road, Oxford OX1 3QG, United Kingdom*

Abstract

Predicting the electricity demand of new supermarkets will help with design, planning, and future energy management. Instead of creating complex site-specific thermal engineering models, simplified statistical energy prediction models as we propose can be useful to energy managers. We have designed and implemented a data-driven method to predict the 'electricity daily load profile' (EDLP) for new stores. Our preliminary work exploits a data-set of hourly electricity meter readings for 196 UK supermarkets from 2012 to 2015. Our method combines the most similar stores on a feature space (floor area split by usage such as general merchandise, food retail and offices and geographical location) to obtain a prediction of the EDLP of a new store. Computational experiments were performed separately for subsets of supermarkets that consume only electricity, both electricity and gas, and by season. The best results were obtained when predicting Summer EDLPs with stores using electricity only. In this case, the average Manhattan difference and the percentage difference are 234 kWh and 16%, respectively. We aim to develop an application tool for supermarket energy managers to automatically generate EDLP for potential new stores.

Keywords: Electricity demand; Prediction; Load profiles; Supermarkets;

* Corresponding author. Tel.: +44(0)1865 610614; fax: : +44(0)1865 610624.
E-mail address: ramon.granell@oerc.ox.ac.uk

1876-6102 © 2019 The Authors. Published by Elsevier Ltd.

10.1016/j.egypro.2019.02.087

Nomenclature

ED	Euclidean distance
EDLP	electricity daily load profile
GM	general merchandising
HVAC	heating, ventilations and air conditioning
*k*NNR	k-Nearest Neighbours Regression Algorithm
MD	Manhattan distance
NP	normalised percentage difference with respect to the original EDLP
SE	stores just with electricity
SEG	stores with electricity and gas
d	number of store features
e_i	electricity consumed (kWh) between the hour i-1 and i
f_i	i-th store feature
k	number of clusters
n	number of stores
F	set of store features
L_s	EDLP of store s
S	set of stores

1. Introduction

The United Kingdom has the target to reduce the greenhouse gas emissions by at least 80%, compared to 1990 levels, by 2050 [1]. The food retail sector accounts for between 3% and 5% of total electricity consumption in UK and 1% of the global CO_2 emissions [2]. Food retail stores are responsible for a significant part of these emissions (it is estimated to be 3-4 % of the electricity production in industrialised economies [3]) as they are energy intensive buildings. In fact, they have the highest energy yearly consumption by area among all type of commercial buildings in the USA [4].

The reasons for this high energy density values are that they include refrigeration, lighting, heating, ventilation and air conditioning (HVAC) of commercial areas which make up the vast magnitude of the total building floor area. In addition, some of them have other utilities such as bakery and catering area.

Analysing and understanding the patterns of energy demand of buildings can help to create measures to reduce this high consumption [5]. In this context, one of the problems that supermarket chains face is that they do not know what is the expected electricity consumption of a new store and if this consumption of the newly opened store is normal compared with similar ones. For this reason, predicting the consumption of the new supermarket can help the company energy manager to plan the energy needs (it is common to create energy consumption budgets significantly in advance). It can also help to detect an anomalous behaviour in terms of consumption of the new store that thus require specific investigation by the energy manager.

There are extensive reviews of methods to predict and benchmark energy use in buildings [6, 7, 8]. However, most of the reviewed works are to predict the electricity for domestic buildings. Works that perform energy performance over non-domestic buildings is the specific topic of [9], but they include over all institutional or public buildings, not commercial buildings. Independently of the type of building, the methodologies that measure and model the energy use of buildings can be classified in two categories: model-driven approaches and data-driven approaches. In model-driven approaches, a high-resolution engineering model of the building that simulates its energy and thermal behaviour is defined. In data-driven approaches, the analysis between the energy performance of the building and the dependent factors are directly modelled with numeric methods such as regression models. They are top-down methodologies as specific building consumption is estimated from large data-sets of several buildings or long series of the aggregated consumption of a specific building. Meanwhile, model-driven approaches can be more accurate than data-driven approaches as they require very detailed level of description of the features of building that is complicated to obtain. For this reason model-driven studies compute their results over just few buildings and data-driven are more easy to test over large set of buildings.

Model driven simulation tool-kits that are currently in use for food retail stores are Supersim [10], EnergyPlus [11] and CyberMart [12]. Energy use measuring and modelling of one frozen food supermarket in London (UK) was proposed by [13]. The model proposed is a calibration of the EnergyPlus tool-kit [12] that coupled refrigeration, HVAC and building systems. They evaluated their results with the energy data available for two years and sp monitoring (CO_2, temperature, lighting level) for one year. There exist more than 100 software tools that simulate and/or model energy consumption in buildings, a directory with most of them can be found in [14].

One of the pioneering works that uses a data-driven approach over smart meter data in supermarket is [15]. The hourly and daily energy of one grocery store in Texas is predicted using 15-min electricity readings during one year. A more recent study that predict the electricity and gas consumption of one supermarket in the UK given the temperature and humidity values is [16]. In this case they predict the weekly consumption (they summed the hourly readings) that is expected from period 2030-2059 considering climate change. A larger data-set of 215 UK hypermarkets were used to estimate the energy consumption with regression models [17]. A model to disaggregate store level energy into weather-dependent and weather-independent components are proposed in [18]. They performed experiments over 94 stores from supermarket chain in World locations with cold winters.

We propose a data-driven model to predict the 'electricity daily load profile' (EDLP) of a new supermarket. Our approach combines EDLPs of similar stores that are from the same supermarket chain whose consumption we aim to predict. Experiments are performed over EDLPs that are obtained from electricity readings for 196 UK supermarkets from 2012 to 2015. The parameters used to predict this consumption are floor area split by usage and geographically location.

The paper is structured in the following way. The proposed method to predict the EDLPs is explained in Section 2. Then we describe the data-set of supermarkets that we use to perform experiments in Section 3. The results of the computational experiments are in Section 4. In the final section we draw some conclusions and propose some lines of future works.

2. Data-driven method to predict the electricity profile

We use EDLPs that are obtained from smart-meter electricity readings with one hour resolution. Each EDLP is the 24-h electricity curve computed when averaging the readings during a specific time period (*e.g.* all days in Winters, weekdays of a whole year). We will have n EDLPs, each one for a particular supermarket that can be analytically seen as a 24 dimensional vector: $L = \{e_1, \ldots, e_{24}\}$ where e_i is the electricity consumed (kWh) between the hour $i-1$ and i. Our method to predict the EDLP is a modification of the k-Nearest Neighbours Regression Algorithm (*k*NNR) [19].

This method is considered as a simple Machine Learning (ML) algorithm that works efficiently when the predicted value can be locally approximated [20]. This local principle (similar 'recipes' yield similar outcomes) seems intuitively correct for the problem that we want to solve. Our hypothesis is that similar supermarkets should show similar patterns of electricity consumption as they have similar business model. Evaluation of the quality of the results and robustness of the algorithm will give us an objective measure. The method produces an easy and fast way to predict the complete load profile. Other ML methods need to be re-adapted to be able to cope with daily profiles.

The proposed algorithm combines the k stores that are most similar to the new one based on a set of building features $F = \{f_1, \ldots, f_d\}$ and has the following steps:

1. Select k and a set of features F.
2. Compute the distance measure from the new supermarket to all the known supermarket using the store features F.
3. Order the known supermarkets by increasing distance and select the k nearest ones.
4. Calculate the value of the new object combining the profiles of the k-nearest neighbours (supermarkets).
5. Compute the error between the real and predicted EDLPs.
6. Repeat the experiment for each new store using leaving-one-out experiments and compute the average error over all the new stores.
7. Repeat the experiment for each combination of (k, F) to find the best combination $(\hat{k}, \hat{F})$.

The search of the best combination of (k, F) can formally expressed by the following equation:

$$(\hat{k},\hat{F}) = \underset{k,F}{\operatorname{argmin}} \sum_{s\in S} Ev\big(L_s, L'_s(k,\mathrm{F})\big) \tag{1}$$

where S is the set of stores, L_s is the real EDLP of store s, $L'_s(k, F)$ is the predicted energy profile when using (k, F) and $\mathrm{Ev}(L_s, L'_s(k, F))$ is the evaluator that measure the error between the predicted and real profile (step 5 of the algorithm). We can use three different evaluators (Ev) to compare the real profile $L_s = \{e_1, \ldots, e_{24}\}$ and the predicted $L'_s = \{e'_1, \ldots, e'_{24}\}$:

Euclidean distance (ED): this is the most common way to compare distances between vectors. Discrepancies between the EDLPs values are accumulated not cancelling between positive and negative values. The distance unit is kWh.

$$\sqrt{\sum_{i=1}^{24}\left(e_i - e'_i\right)^2} \tag{2}$$

Manhattan distance (MD): similar as before but it is easier to understand as it is just the addition of differences in absolute value. The distance unit is again kWh. As it is used absolute value, both positive and negative errors accumulated.

$$\sum_{i=1}^{24}(|e_i - e'_i|) \tag{3}$$

Normalised percentage difference with respect to the original EDLP (NP): same as MD but normalising with the total consumption of the original EDLP. This relative distance considers the proportion of the error with respect to the total consumption of the original profile, i.e. is not the same to have 100 kWh of error when the original profile uses a total of 1000 kWh or 10000 kWh. A value of 100% will indicate that the additions of errors is the same that the total real energy consumed.

$$\frac{100 * \sum_{i=1}^{24}|e_i - e'_i|}{\sum_{i=1}^{24} e_i} \tag{4}$$

Independently of the evaluator, there should be a way to extend it to summarize the predicted error over all the new stores in the data-set (step 6 of the algorithm). We compute the mean over all the predicted EDLPs for the three distances previously described.

From the steps of previous algorithm, one of the main concepts that needs to be properly defined is how to describe the similarity between stores in the store features domain F, *i.e.* how to compute the distance measure between two stores based on F (step 2 of the algorithm). These features F include the floor area split by usage and the geographical location of the store (detailed description is given in Section 3) and the distance is the Euclidean distance over all these features. The step 3 of the algorithm needs also to be clarified, we need to predict the ELDP L, of the new store using the ELDPs of the k closest supermarkets. We just compute the mean over the EDLPs of these supermarkets *i.e.* the mean over the k vectors L dimension by dimension.

In our algorithm there are two parameters that need to be estimated: the number of nearest neighbours k and the final configuration of features F. As for each store s with real EDLP L_s, we have previously described how to compute the predicted energy profile L,s combining the k-NN profiles of the stores based on features F, we have just to find the combination of k and F that minimizes the total prediction error over the evaluators (Equation 1). For each possible combinations of parameters (k, F), we perform all the experiments and compare the predicted EDLPs with the real EDLPs and select the parameters that minimises the ED evaluator. This brute-force search to find ($\hat{k}$, $\hat{F}$) is performed using $n - 1$ stores. This leaving-one-out technique is a common ML experimental set-up [20] in which all the data points except the one being estimated are used as predictors. Then the same experiment are repeated n times selecting each time a different point to predict. In our case this can be seen as assuming that each time a different supermarket of our data-set is the new one whose EDLP needs to be predicted using the EDLPs of the n 1 other supermarkets. The search space for k goes from 1 to 50 and F comprises all the different feature combinations but not the empty one: $2^d - 1$ different combinations. In our case as we use seven features, we have 127 combinations.

3. Data-set of energy readings of supermarkets

The data-set used in this work is formed by electricity readings of 196 supermarkets of the same retailer distributed geographically across the UK. Hourly resolution readings are able from January 2012 to December 2015, however not all the stores have all values as some opened later than January 2012, closed before December 2015 or have missing/erroneous values. Readings are considered valid if they have a value greater than zero and lower than 400 kWh (1.47% of the readings were removed). After this filter, EDLPs are computed with all the available valid readings for each store for Monday-Saturdays (Sundays have a different profile as store opening hours are characteristically different).

As meta-data we have the following features: floor area divided by usage (see Table 1), geographical coordinates and if they use gas or not. This last feature is used to create two separated subsets based on the energy they use: stores just with electricity (SE) and stores with electricity and gas (SEG). Electricity consumption values and patterns are affected if the store uses gas for heating and other services. As an example of this difference we show the Winter EDLPs normalised by sales area of the stores of the subsets SE and SEG in Figure 1. We displayed normalised consumption by area since stores can have very different size as the min, max and SD values of Table 1 indicate. There are 86 (43.9%) supermarkets that use just electricity and and 110 (56.1%) supermarkets that use electricity and gas. The rest of features (six floor area features in Table 1) and geographical location are used to find the closest supermarkets in our algorithm (feature F in Equation 1).

Table 1. Floor features and values over the supermarkets. Sales area is the addition of GM, Food and Cafeteria areas.

Area	Min (m^2)	Max (m^2)	Avg (m^2)	SD (m^2)
General merchandising (GM)	0.0	648.2	54.0	92.7
Food	159.3	1590.3	644.0	224.6
Cafeteria	0.0	269.4	41.9	59.6
Sales	164.0	1925.7	739.9	292.3
Office	0.0	540.7	160.0	91.3
Total	183.5	2305.8	899.8	358.2

4. Results

Computational experiments are performed separately on the two subsets of supermarkets (SE and SEG). For each of the supermarkets of these two subsets, three seasonal EDLPs are independently computed over all of the available readings: Winter (December, January and February), Summer (June, July and August) and Spring/Autumn (March, April, May, September, October, November). With this season separation, we partially take into account the different patterns of consumptions due to the external meteorological conditions. Code that computes the method and performs all the computational experiments was implemented in C++.

The results for the three prediction evaluators over all the experiments are in Table 4. Each row of this table are the results obtained with the best combination of features and number of clusters $(\hat{k}, \hat{F})$. For instance when predicting Winter profiles of SE the best configuration is $\hat{k}$= 6 and $\hat{F}$={GM area, Food area, Cafeteria area}. By average, ND for Winter profiles of stores just with electricity is 18%. However, there are several stores whose EDLP were better predicted. The ND for the best feature combination for these stores is displayed in Figure 2. The 50% of the store has a ND lower than 15.6% of error. There is 10% of the stores whose NDs are greater than 30%. These outliers with so large error are stores whose features are not similar to most of the other stores. The GM area of these stores is unusually large (more than 500 m^2) when the average is 54 m^2 and just seven stores have more than 300 m^2.

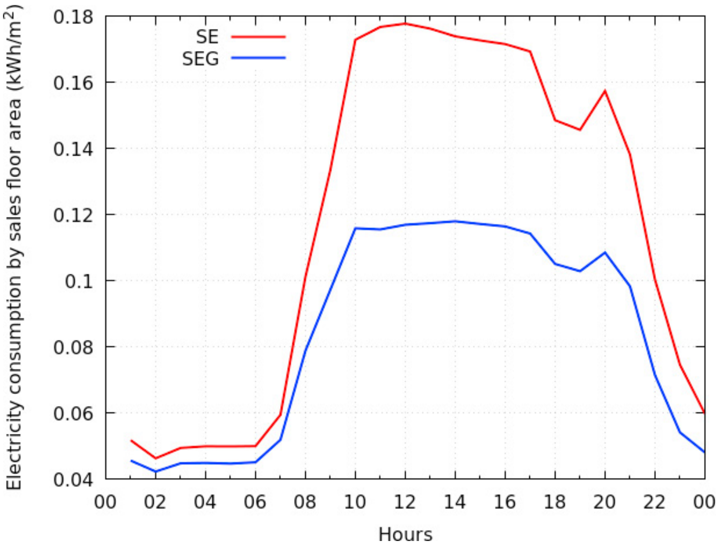

Fig. 1. Average of all the Winter electricity by sales area profiles of the two subsets of supermarkets (SE are supermarkets just with electricity and SEG are stores with electricity and gas).

Stores using only electricity have lower error than stores with electricity and gas. In addition Winter errors are higher than Summer and Spring/Autumn errors. These two facts may imply that heating is increasing the prediction error and need to be better considered. Stores just with gas should be using gas heating systems, however it is possible that some of them use also electricity heating. In general, we can think that electricity consumption used by the heating system may be quite different from one store to another, making the predictions in Winter more difficult than during the other seasons. This is almost certain because of distinct building techniques, materials and standards for different ages of buildings.

As examples of the predicted and real EDLPs we display the EDLPs of the stores whose ND are in the median for Winter and Summer in Figure 3. These profiles are examples of what may be considered average errors in the predictions. The ED, MD and ND for the Winter profile of store one are 63.1 kWh, 280.3 kWh and 15.6% respectively. The ED, MD and ND for the Summer profile of store two are 32.8 kWh, 134.2 kWh and 12.1% respectively. Summer profiles are obtained with $\hat{k}$ = 5 and $\hat{F}$ = {Food area, Cafeteria area, Sales area, Office area ,Total area}

Table 2. Average evaluators when predicting the EDLPs. ED is Euclidean distance, MD is the Manhattan distance and ND is the normalised percentage difference with respect to the original EDLP

	Stores just with electricity			Stores with electricity and gas Season		
(%)	ED (kWh)	MD (kWh)	ND (%)	ED(kWh)	MD (kWh)	ND
Winter	72	295	18	121	491	22
Summer	54	234	16	64	262	15
Spring/Autumn	59	241	16	86	347	17

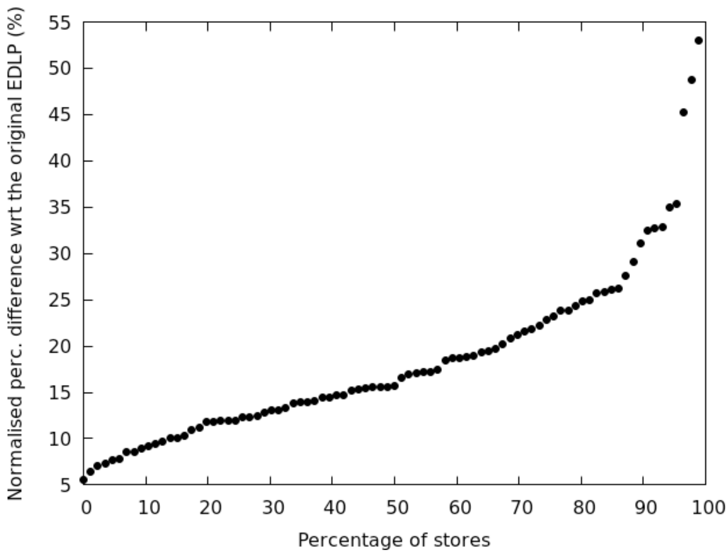

Fig. 2. Normalised percentage difference with respect to the original EDLP when predicting Winter EDLP of stores just with electricity.

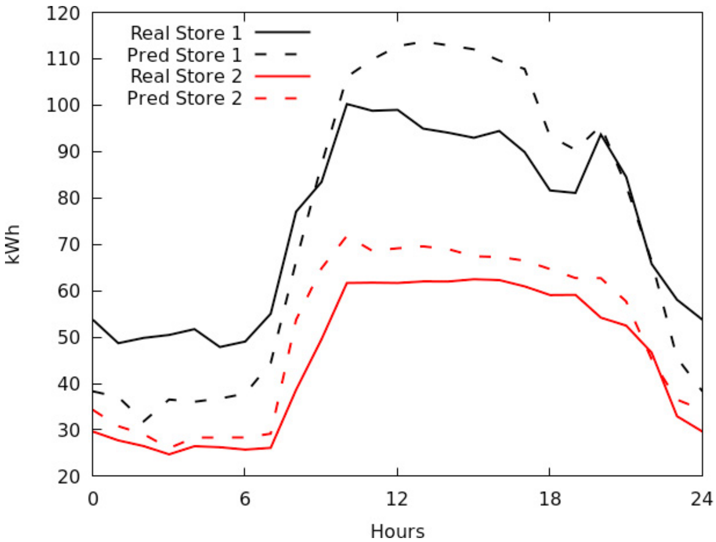

Fig. 3. Predicted and real profiles of the stores with the median error when predicting Winter and Summer EDLP of stores just with electricity.

5. Conclusion and future work

We have presented a generic method to predict the EDLPs of new stores that can be used on any data-set when both electricity readings and store features are available. Prediction experiments over a data-set of 196 UK supermarkets were performed showing that our method can predict the profile on new stores with average error of between 15% and 22% depending on the season and fuel consumed by the stores. Generally, EDLPs of stores with building features similar to existing ones are better estimated than those with abnormal features such as very large general merchandise sales areas.

There are several research lines that follow from this work. We can use more complex methods to combine the

EDLPs of the k nearest neighbours to compute the predicted profile. Currently, we gave the same weight to all neighbours. There are other options in which smaller weight is assigned for points farther from the predicted point, existing different ways to compute them: a weighted average [19] or a kernel function [21]. Another interesting option is the the use of confidence intervals when predicting the profile instead of giving an exact value.

Potentially, temperature (meteorological) information can be incorporated using heating and cooling degree days [22]. Furthermore, we will examine how features such as age, construction standards, and materials can be used to reduce errors.

Acknowledgements

We want to thanks the supermarket company to provide the data used in this work. This work was partially funded by EPSRC Impact Acceleration Account - University of Oxford EP/R511742/1 (R. Granell and D.C.H Wallom).

References

[1] Energy and Climate Change Committee, Building a low-carbon economy the UKs contribution to tackling climate change (2009).
[2] Tassou S A, Ge Y, Hadawey A, Marriott D, Energy consumption and conservation in food retailing, Applied Thermal Engineering 31 (2-3) (2010) 147.
URL https://hal.archives-ouvertes.fr/hal-00692330
[3] Kauko H, Husevåg Kvalsvik K, Masson N, Noel C, Minetto S, Rossetti A, Marinetti S, Thalheim D, Martens, Karampour K, Pis- copiello S, Fidorra N, Gimeno Frontera B, Saez de Guinoa A, Mainar Toledo L, Ciconkov S, Ciconkov V, Development of the EU ecolabel criteria for food retail stores : Preliminary report, Tech. rep., KTH, Applied Thermodynamics and Refrigeration (2017).
[4] Energy Star, U.S. energy use intensity by property type, Tech. rep., Energy Star (2016).
URL https://portfoliomanager.energystar.gov/pdf/reference/US%20National%20Median%20Table.pdf
[5] Centre for Sustainable Energy, Generating value from smart meter data.
URL https://www.cse.org.uk/projects/view/1210
[6] Chung W, Review of building energy-use performance benchmarking methodologies, Applied Energy 88 (5) (2011) 1470 –1479.
URL http://www.sciencedirect.com/science/article/pii/S030626191000485X
[7] Li Z, Han Y, Xu P, Methods for benchmarking building energy consumption against its past or intended performance: An overview, Applied Energy 124 (2014) 325 – 334.
URL http://www.sciencedirect.com/science/article/pii/S0306261914002505
[8] Zhao H, Magouls F, A review on the prediction of building energy consumption, Renewable and Sustainable Energy Reviews 16 (6) (2012) 3586 – 3592.
URL http://www.sciencedirect.com/science/article/pii/S1364032112001438
[9] Borgstein E, Lamberts R, Hensen J, Evaluating energy performance in non-domestic buildings: A review, Energy and Buildings 128 (2016) 734 – 755.
URL http://www.sciencedirect.com/science/article/pii/S0378778816306144
[10] Arias J, Lundqvist P, Modelling and experimental validation of advanced refrigeration systems in supermarkets, Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering 219 (2) (2005) 149–156.
URL http://dx.doi.org/10.1243/095440805X8593
[11] US Department of Energy Building Technologies Office, EnergyPlus simulation software, accessed: 24-04-2017.
URL www.eere.energy.gov/buildings/energyplus
[12] Ge Y, Tassou S A, Mathematical modelling of supermarket refrigeration systems for design, energy prediction and control, Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy 214 (2) (2000) 101–114.
[13] Mylona Z, Kolokotroni M, Tassou S A, Frozen food retail: Measuring and modelling energy use and space environmental systems in an operational supermarket, Energy and Buildings 144 (2017) 129 – 143.
URL http://www.sciencedirect.com/science/article/pii/S0378778817309544
[14] International Building Performance Simulation Association, Building energy software tools, accessed: 24-04-2017.
URL http://www.buildingenergysoftwaretools.com
[15] Schrock D W, Clarige D E, Predicting energy usage in a supermarket, in: Proceedings of the Sixth Symposium on Improving Building Systems in Hot and Humid Climates, 1989, pp. 19–27.
[16] Braun M, Altan H, Beck S, Using regression analysis to predict the future energy consumption of a supermarket in the UK, Applied Energy 130 (2014) 305 – 313.
URL http://www.sciencedirect.com/science/article/pii/S0306261914005674
[17] Spyrou M S, Shanks K, Cook M J, Pitcher J, Lee R, An empirical study of electricity and gas demand drivers in large food retail buildings of a national organisation, Energy and Buildings 68, Part A (2014) 172 – 182.
URL http://www.sciencedirect.com/science/article/pii/S0378778813005914
[18] Iyer S R, Sankar M, Ramakrishna P V, Sarangan V, Vasan A, Sivasubramaniam A, Energy disaggregation analysis of a supermarket chain using a facility-model, Energy and Buildings 97 (2015) 65 – 76.

URL http://www.sciencedirect.com/science/article/pii/S0378778815002716
[19] Altman N S, An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression, The American Statistician 46 (3) (1992) 175–185.
[20] Bishop C M, Pattern Recognition and Machine Learning, Springer-Verlag, 2006.
[21] Wand M, Jones M, Kernel Smoothing, Chapman & Hall/CRC Monographs on Statistics & Applied Probability, Taylor & Francis, 1994.
[22] Day T, Degree-days: theory and application, The Chartered Institution of Building Services Engineers (2006).