Highlights

Predicting electricity demand profiles of new supermarkets using machine learning

Ramon Granell, Colin J. Axon, Maria Kolokotroni, David C.H. Wallom

- We compare 4 ML techniques to predict daily electricity load profiles for supermarkets
- Model inputs include floor area and use type, and location
- ML prediction used 1-h-resolution electricity readings of 213 UK supermarkets for 6 years
- Best prediction results are 12% average errors for 2017 Summer profiles
- Profiles computed for warm periods are predicted better than for cold periods

Predicting electricity demand profiles of new supermarkets using machine learning

Ramon Granell[a,b,1], Colin J. Axon[a], Maria Kolokotroni[a], David C.H. Wallom[b]

[a] *Institute of Energy Futures, Brunel University London, Uxbridge, London 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 consumption of proposed new supermarkets is helpful to design and plan future energy management. Instead of creating complex site-specific thermal engineering models, data-driven energy prediction models can be useful to energy managers. We have designed and implemented a data-driven method to predict the future 'electricity daily load profile' (EDLP) of new supermarkets using historical EDLPs of existing supermarkets of the same type. The supermarket features used for the prediction are 10 types of floor areas divided by usage (m^2) and its location. Four data-driven regression models are used and compared to predict EDLPs: Artificial Neural Networks, Support Vector Machines, k-Nearest Neighbours and OLS. Prediction computational experiments were performed over 1-h electricity readings of 213 UK supermarkets gathered during six years. Prediction error mainly varies between 12 and 20% depending on method, year, supermarket type, and division of the data (season or temperature intervals). EDLPs computed over warm periods are better predicted than over cold periods and supermarkets only with electricity are better predicted than supermarkets with electricity and gas. The three features with more weight in the prediction are Food, Chilled produce and Cafeteria areas. The limitations of machine learning methods to solve this problem are discussed.

Keywords: electricity use profile, commercial, prediction, supermarket, energy analytics

Email address: ramon.granell@oerc.ox.ac.uk (Ramon Granell)

[1] Corresponding author.

Glossary

ANN Artificial neural networks

ED Euclidean distance

EDLP Electricity daily load profile

GHG Greenhouse gas

HVAC Heating, ventilation and air conditioning

kNNR k-nearest neighbours regression

ML Machine learning

MSE Mean squared error

NP Normalised percentage difference with respect to the original EDLP

OLS Ordinary least of squares

SE Supermarkets using only electricity

SEG Supermarkets using electricity and gas

SVR Support vector regression

Symbols

e_i electricity consumed (kWh) between the $(i-1)$-th and i-th time interval

k number of EDLPs used for the prediction

p number of previous years used to predict the EDLP

y year used to compute the EDLP

D number of time intervals of the EDLP

F set of supermarket features used to predict the EDLP

L_s EDLP of the supermarket s

S, S' sets of new and existing supermarkets respectively

1. Introduction

Many governments, including those of EU countries [?], have committed to reducing greenhouse gas (GHG) emissions to net zero. The UK [?] aims to achieve this by 2050 using a series of carbon budgets [?]. Energy use in buildings accounts for more than 30% of global final energy demand [?] 36% of GHG emissions in EU countries [?], with the UK building stock accounting for 88 $MtCO_2$ in 2018 [?]. Therefore, reducing energy use by all building types (residential, commercial, services and industrial) is one of the targets [?], with demand pattern analysis [?] and the use of building codes [?] helping create efficiency measures to lower consumption [?]. Much attention has been paid to the residential sector [? ? ?], whilst consumption in commercial and industrial buildings has been under-investigated because of its diversity, lack of publicly access data, and the nature of property ownership [? ?]. Globally, the commercial and public sector consumed 8% of the total energy in 2018 (industry: 38%, transport: 29%, residential: 21% and others: 4%) [?]. However, czonsidering only electricity, commercial sector accounts 21% of the total final consumption. Food retail stores (supermarkets) consume 3-4% of the electricity in industrialised economies [?]. Moreover, supermarkets are among the type of commercial buildings with the highest consumption by floor area [?]. The main demands for energy are refrigeration, heating, ventilation and air conditioning (HVAC), and lighting which make up the majority of the building's floor area. Furthermore, some supermarkets have facilities such as a bakery, hot food preparation, or a cafeteria.

In managing a portfolio of stores (food or other retail) total energy demand and the temporal profile are useful performance indicators, though inevitably there will be differences between stores. The differences arise due to building attributes, *e.g.* age, size, levels of insulation and construction type; in-store facilities and appliances, *e.g.* technologies used and their age, and maintenance; as well as usage patterns and geographical location. Knowing the expected demand of a supermarket informs energy management decisions and establishes a baseline for measures to reduce the consumption, *e.g.* supermarkets with the expected higher consumption can be the priority to implement these measures. The interest in estimating electricity consumption is threefold: 1) planning the annual electricity budget for the portfolio of stores, 2) negotiating energy supply contracts, and 3) detecting supermarkets with unexpected discrepancy between the estimated and actual demand (given a robust prediction method).

But what constitutes normal use of a particular supermarket or reasonable use for stores with similar characteristics? Furthermore, when a company considers adding a supermarket to its portfolio, what will be a reasonable amount of energy for it to use [**?**]? If a store is using more energy than expected, investigations can be made with the potential of interventions to mitigate the additional energy use. A large portfolio of sites may render manual monitoring too expensive or difficult, thus we examine 'what is normal' in an automated manner. We focus on predicting the typical electricity daily load profile (EDLP) of a new supermarket. Altthough both gas and electricity are used at present, there is a trend to replace gas with electrical heating in the UK [**?**] to help reduce the CO_2/kWh intensity.

We explore four machine learning (ML) methods to predict EDLPs of supermarkets using hourly electricity data during a period of six years. The data-set was obtained from a portfolio of a UK supermarket chain with 213 supermarkets. The main questions that we try to answer are:

- can the EDLP of new supermarkets be predicted accurately using the proposed ML algorithms, and what are the most suitable metrics for evaluating the quality of the predicted EDLP?
- how much data is enough, and which subset of readings are the most useful — year, season or a temperature criterion?
- which supermarket or building features are more helpful? This can guide decisions on which attributes should be monitored.

The paper has the following structure. We review the literature of previous studies in Section 2. The techniques used for predicting the EDLPs of new supermarkets and the data-sets are described in Section 3.1. Results and discussion are presented in Section 4. Finally, conclusions are drawn in Section 5 where we suggest possible future lines of research.

2. Background

Independently of the type of building, the principle prediction methods can be divided into two approaches, namely, model-driven and data-based. The model-driven approach uses sophisticated high-resolution engineering methods based on the thermal, energy and architectural features of the building to simulate its future energy behaviour. In data-driven approaches, the energy performance of the building is directly modelled with numerical and statistical methods. As input to the prediction method for data-driven models, only some general features obtained from data-sets are used *e.g.* temperature series, electricity readings. Model-driven approaches are usually more accurate than data-driven approaches, however, they are more complicated and computationally expensive. Thus model-driven studies compute their results for specific buildings while data-driven models can be used large sets of buildings. There are extensive reviews on methods to predict and benchmark energy use in buildings [? ? ? ?], however, most of the reviewed works predict electricity of dwellings or offices. Previous investigations have not yielded robust methods for predicting energy of targeted types of non-domestic buildings for retail use.

One of the pioneering data-driven supermarket studies [?] used a year of 15-min electricity readings of one grocery store in Texas to predict hourly and daily consumption using a change-point algorithm. A more recent study [?] predicted weekly (aggregated hourly readings) electricity and gas consumption for one UK supermarket using temperature and humidity values, and projected for the period 2030-2059 to consider climate change. A larger data-set of 215 UK hypermarkets were used to estimate the total annual electricity demand with linear regression models [?]. In [?], the annual energy-use intensity is estimated for 30 supermarkets with a linear regression model having as input building features such as floor area and building age and other features, *e.g.* number of customers. Electricity consumed by the HVAC and refrigeration systems of one supermarket is predicted using Artificial Neural Networks (ANN) by [?].

There are several reviews [? ?] using ML techniques to predict electricity demand in all types of buildings, but we focus on the techniques we exploit in this study. Artificial Neural Networks (ANN) have been used to predict the annual and monthly heating demand of small Swedish domestic buildings [?] and HVAC loads in a Spanish hotel [?]. Support Vector regression (SVR) models were used by [?] to predict monthly energy consumption of four commercial building in Singapore. Models based on SVR have also been used to predict the energy load (hours to days) of a French residential building [?]. ANN and SVR were compared when predicting hourly cooling load in an office building in China [?] and hourly energy consumption of an office building in Shanghai [?]. The k-Nearest Neighbour (kNN) algorithm was used to forecast the next day consumption of 6,000 domestic Irish buildings in [?], and for the hourly air conditioning load of an office building in

China [?].

We note several gaps in the literature, which we aim to address. First, we will characterise the important similarities between multiple stores and use historical data to predict consumption for an unknown (new) supermarket. This is different from the two most typical approaches in the literature where: 1) future consumption of the same supermarket is predicted using its historical data *e.g.* [?] or 2) consumption of unknown supermarkets is predicted using other supermaket consumption during the same time period (not historical) *e.g.* [?]. Secondly, we will use daily profiles and account for seasonal variations in consumption, instead of a unique aggregated daily, weekly, monthly or annual value *e.g.* [? ? ?]. And thirdly, we compare four different ML methods across a five-year span of data for 213 supermarkets, more than previous works that predict electricity use in supermarkets. These represent novel contributions to the knoweledge-base of energy use in supermarkets.

3. Methods

First we will state the problem in a formal manner, then describe the data-set and its preparation, and finally introduce the ML techniques and their implementation.

Formally, the problem is defined as predicting the daily profile $L_s = e_1, \ldots, e_D$ of a new supermarket $s \in S$ for a year y based on historical profiles of existing supermarkets S' and the supermarket features F. L_s is the EDLP of the new supermarket s, e_i is the electricity consumed (kWh) between the $(i-1)$-th and i-th time interval, D is the number of intervals, S and S' are the set of new and existing historical supermarkets, respectively ($S \cap S' = \emptyset$). The features F is the set of available information about the supermarket building such as the floor area divided by usage and the supermarket geographical location. Independently of the particular prediction method to use, the experimental framework is the following:

1. Select set of features (F) and number of supermarkets used to predict (k)
2. Predict the EDLP L_s using historical EDLPs of existing k similar supermarkets
3. Compute the error between the real and predicted EDLPs.
4. Repeat the steps 2 and 3 for each new supermarket $s \in S$.

5. Repeat the experiments (steps 1-4) for each combination of (k, F) to find the best combination $(\hat{k}, \hat{F})$.

Step one of the algorithm determines the selection of the features (F) and number of EDLPs (k) to be used for the prediction. They are the global parameters of the model. The search of the best combination of (k, F) (step 5) can formally expressed by Equation 1.

$$(\hat{k}, \hat{F}) = \underset{k,F}{\arg\min} \sum_{s \in S} \mathrm{Ev}(L_s, L_s(k, F))) \tag{1}$$

where S is the set of new supermarkets, L_s is the real EDLP of supermarket s, $L_s(k, F)$ is the predicted energy profile when using parameters (k, F) and $\mathrm{Ev}(L_s, L'_s(k, F))$ is the evaluator that measures the error between the predicted and real profile (step three of the algorithm).

Step two of the algorithm depends on the prediction method. Comparing the results obtained by different algorithms will provide us a reference of the difficulty of the stated problem.

3.1. The Data-set

The data-set comprises 1-h resolution electricity meter readings (kWh) from 213 UK supermarkets of the same chain for the period 2012–17. The meta-data features available of each supermarket are:

Floor area: subdivided into 8 use-categories (m^2): General Merchandising (GM), Food, Cafeteria, Office, Storage, Chilled, Frozen, and Produce. The Total area is also given, and the Sales area is the sum of the GM, Food and Cafeteria areas. Data on the Chilled, Frozen and Produce areas was available for only five supermarkets. For the other supermarkets, these three categories were estimated with a linear regression model, using the other areas as predictors. These 10 features F (Table 1) are used as input to the prediction models.

Geographical location. longitude and latitude.

Temperature readings: daily average external temperature values (°C) provided by the company are available for all days of 2015–17.

Fuels types: there are supermarkets that use electricity and gas (SEG) and others use only electricity (SE). The computed profile only considers those supplied by electricity alone but experiments are always performed independently over the SEG and SE data-sets.

Area Type	Min (m^2)	Max (m^2)	Avg (m^2)	SD (m^2)
Total	324.6	3279.3	1242.7	471.6
GM	1.4	572.8	47.9	78.5
Food	162.1	1590.3	700.8	248.2
Cafeteria	0.0	269.4	39.0	58.5
Sales	164.0	1925.7	787.6	312.9
Office	0.0	540.7	157.5	88.2
Storage	0.0	973.5	297.7	136.1
Chilled	22.2	38.9	28.5	2.9
Frozen	0.3	4.8	2.0	0.8
Produce	0.0	12.3	3.1	2.3

Table 1: Floor features and values for the supermarket set.

The electricity readings are divided temporally to compute the EDLPs based on various criteria. First, they are divided by years as the goal is to predict the consumption of new supermarkets for the coming year. As readings are available from 2012-2017, daily profiles of new supermarkets of each individual year from 2013 to 2017 are predicted using historical data. Generically, if an EDLP of year y is predicted for one supermarket, profiles of other supermarkets computed with readings from previous years: from years $y-p$ to $y-1$, can be used. This window width p is also a parameter for the experiments as we do not know how many years of historical data to use to predict future profiles of new supermarkets more accurately. Secondly, only the Monday to Saturday readings are selected, because Sunday opening and closing times vary widely. In addition to these two temporal divisions, two sets of experiments based on weather conditions are investigated:

Seasons UK meteorological conditions vary widely, affecting energy consumption likewise. Three seasonal EDLPs are independently computed over all available readings of the selected year: Winter (December, January and February), Summer (June, July and August) and Spring/Autumn (March, April, May, September, October, November). Fig. 1 shows the profiles for the SE and SEG groups computed over the Winter, Summer, Spring/Autumn 2017 readings. The seasonal differences are more important for SE group as electricity is used for heating. Experiments predicting EDLPs computed over all the available years

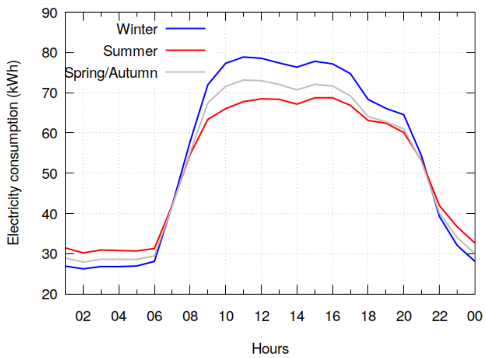
($y \in [2013, 2017]$) and possible values for parameter p ($p = 1, \ldots, 5$ when $y - p \geq 2012$) are performed. An independent prediction experiment is performed for each year y, window width p and season. Table 2 shows the number of supermarkets for testing (number of supermarkets with readings in year y) and training (number of supermarkets with enough readings in years $y - p$ to $y - 1$) the ML algorithms.

Temperature The external temperature data allows us to split the days during 2015-17 based on the average daily temperature. Days are divided using temperature intervals of 1 °C, but larger intervals are allowed in the extremes as there are insufficient supermarkets with readings during days with extreme temperatures. For each temperature interval, the EDLP of each supermarket is computed using only the days that have the temperature in the interval, *i.e.* it is treated as an independent prediction problem. For these experiments, only the 2017 EDLPs are predicted using EDLPs computed with 2015-16 readings. This is done because a sufficient number of days with readings for each temperature interval exist, though not all supermarkets have days with readings for all intervals (at the low/high extremes). For the coolest and hottest temperatures, all days are grouped as ≤ -3°C and > 23°C intervals respectively. There is a total of 28 different temperature intervals. For the 21 temperature intervals between $]-1, 0]$ to $]19, 20]$ °C there are available data in more than 95% of the supermarkets for both the SE and SEG groups (84 and 129, respectively). In the extreme intervals, there are fewer supermarkets with available readings of days with these temperatures. Intervals with days ≤ -3°C and $]-3, 2]$°C contain fewer than 30% of the total supermarkets.

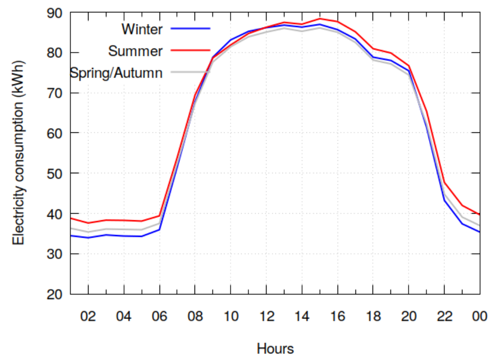
3.2. Machine Learning Techniques and Computational Experiments

We exploit four different approaches based on established ML techniques, each of different mathematical nature. All these techniques have been used in isolation for predicting electricity consumption.

kNNR the k-Nearest Neighbours Regression Algorithm (kNNR) [**?**] is considered as a simple and fast ML algorithm that works efficiently when the predicted value can be locally approximated [**?**]. In our case, the hypothesis is that similar supermarkets should show similar patterns of electricity consumption. The method predicts the complete



(a) Supermarkets with electricity only



(b) Supermarkets with electricity and gas

Figure 1: Seasonal electricity profiles of all the supermarkets during 2017.

load profile, combining the k supermarkets that are most similar to the new one based on a set of supermarket features F. To find the most similar supermarkets we compute the distance of the new supermarket with the complete set over features F. Due to the dual nature of the features (location and floor area), they are individually computed using Euclidean distance, normalised and finally averaged. Later, the k supermarkets with lowest distance are selected and their EDLPs averaged to compute $L_s(k, F)$

$$e'_i = \frac{\sum_{s \in S_k} e_{i,s}}{k}, \;\; 1 \leq i \leq D \tag{2}$$

where e'_i is the predicted electricity value at i-th time, $e_{i,s}$ is the real historical electricity value at i-th time of the s supermarket and S_k is the set with the k most similar supermarkets to the one to predict. We have also implemented variations of Equation 2 in which a weighted averaged is computed based on a kernel weighted function, *e.g.* Epanechnikov Quadratic equation and Tri-cube function.

OLS the ordinary least of squares (OLS) [**?**] is a linear regression model that estimates the unknown parameters minimising the sum of squares of residuals. Under the assumptions that the model parameters must be linear and that the residuals are normally distributed, the OLS parameters are estimated with the Maximum Likelihood approach. In our case, each data-point of the EDLP ($e_i, 1 \leq i \leq D$) is individually computed following Equation 3 using the same parameters: the k

	Pred. year		Previous years used to train (p)				
	Year (y)	#Test	One	Two	Three	Four	Five
SE	2017	84	84	84	84	84	85
	2016	84	83	83	83	84	-
	2015	83	81	81	82	-	-
	2014	81	81	82	-	-	-
	2013	81	81	-	-	-	-
SEG	2017	129	129	129	129	129	129
	2016	129	111	111	111	111	-
	2015	111	98	98	98	-	-
	2014	98	87	87	-	-	-
	2013	87	78	-	-	-	-

Table 2: Number of supermarkets used for testing and training for the seasonal experiments depending on the historical years used.

closest EDLPs and F as predictors for the regression.

$$e'_i = \beta_{i,1} f_1 + \beta_{i,2} f_2 + \ldots + \beta_{i,|F|} f_{|F|} + \epsilon_i \quad (3)$$

where $\beta_{i,j}$ is the predicted coefficients that multiplies the j-th feature when estimating the i-th electricity reading and ϵ_i is the estimated intercept.

ANN the artificial neural network regression model (ANN) [**?**] are parametric models based on the linear combination of a fixed number of non-linear functions as Equation 4 indicates for one neuron.

$$g(b + \sum_{i=1}^{|F|} f_i w_i) \quad (4)$$

where w_i is the weight for i-th feature f_i, b is the bias and $g()$ is the non-linear function that can be a sigmoid function. We combine neurons into layers and the morphology of the network (number of layers and neurons per layer) is designed based on the number of input features. Then the parameters of the network were computed using the backpropagation algorithm.

SVR a support vector machine regression model (SVR) [?] is a non-probabilistic supervised algorithm. New point estimation depends on the evaluation of kernel function trained with data points (support vectors) that divides the domain space. The generic function to predict a new value is in Equation 5.

$$g(X) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) K(X_i, X) + b \tag{5}$$

where X are the observations (features in our case), N is the number of data points, α_i, α_i* and b are estimated model parameters and $K()$ is the kernel function *e.g.* linear, polynomial, sigmoid, Radial Basis functions (RBF).

For the OLS, ANN and SVR methods, each point of the EDLPs is individually predicted (*i.e.* different model parameters need to be estimated for each dimension), but the whole EDLP is directly estimated using the kNNR.

We use three evaluators to asses the error between the prediction obtained with one of the previous methods and the real EDLP is computed in step three of the algorithm:

Euclidean Distance (ED) in which discrepancies between the EDLPs absolute values are accumulated (in kWh),

$$\sqrt{\sum_{i=1}^{D} (e_i - e'_i)^2} \tag{6}$$

where e_i and e'_i are respectively the real and estimated consumption value at time i.

Mean Squared Error (MSE) in which absolute values are computed and normalised by the dimension (number of hours) and number of stores.

$$\frac{1}{|S|} \sum_{s \in S} \frac{1}{D} \sum_{i=1}^{D} (e_{i,s} - e'_{i,s})^2 \tag{7}$$

where $e_{i,s}$ and $e'_{i,s}$ are respectively the real and estimated consumption value at time i for store s.

Normalised Percentage (NP) difference with respect to the original EDLP (NP) computes the relative distance considering the proportion of the error with respect to the total consumption of the original profile,

$$\frac{100 * \sum_{i=1}^{D} |e_i - e'_i|}{\sum_{i=1}^{D} e_i} \tag{8}$$

This evaluator has the advantage of capturing the relation of the error considering the total energy consumer.

The ED and NP evaluators are extended to summarize the predicted error over all the set of new supermarkets S. We compute the mean of the evaluator over all the predicted EDLPs, for instance the averaged ED:

$$\overline{ED} = \frac{\sum_{s \in S} ED_s}{|S|} \tag{9}$$

where ED_s is the ED computed over the real and predicted EDLPs of the supermarket s. The evaluator $\overline{ED}$ is also used in Equation 1 to search the combination of k and F that minimizes the total prediction error over all the new supermarkets S. The evaluator $\overline{NP}$ is computed in the same way.

For our case study, approximately 30 new supermarkets are opened each year. To give robust and significant results, we assume that each supermarket is considered a new one and the others $|S| - 1$ are used to predict the EDLPs of the new one. This leaving-one-out technique is a common ML experimental set-up [**?**] for small data-sets in which all the data points except the one being estimated are used as predictors. Then the same experiment is repeated $|S|$ times selecting each time a different point to predict. The EDLPs computed over historical data (years $y - p, \ldots, y - 1$) are used to compute the EDLP of the new one for year y. More details about how to compute the EDLPs are given in Section 3.1.

Error bars are computed to model the uncertainty of the prediction, *i.e.* predicting an interval instead of a single line of the EDLP is helpful to have a broader estimation of the possible EDLP. They are calculated adding/subtracting twice the value of the standard error computed over the k EDLPs to the predicted value.

For the seasonal and temperature data, each algorithm has parameters and functions to configure. For the kNNR algorithm, in addition to the averaged model (Equation 2), two more sophisticated kernel-weighted functions

(Epanechnikov Quadratic and Tri-cube functions) were also implemented but no improvement was found. For the ANN, we use a logistic function ($g()$ (Equation 4)) over a two internal layers net, *i.e.* the configuration of the network is $|F|$-4-2-1, where $|F|$ is the number of features. For the SVR, a RBF kernel function ($K()$ (Equation 5)) was used as it models non-linearly the input data features to predict.

Independent of the prediction algorithm, we used the brute-force approach (Equation 1) searching all combinations of parameters $(\hat{k}, \hat{F})$. The maximum number of combinations, for each one of the season- and temperature-divided experiments, is $(2^{|F|}-1)*(|S|-1) = (2^{|11|}-1)*(129-1) = 262,016$, and multiplied by $|S|$ for the leaving-one-out approach. Thus for the temporally more complex methods (ANN and SVR) we used stepwise regression [**?**] with the whole feature set F (using all the supermarkets, $k = |S|$). This reduces the combinations to $\sum_{i=1}^{11} i = 66$. For the OLS, we also used stepwise regression but scanning over all the values of k: $\sum_{i=1}^{11} i*(129-1) = 8,448$ combinations. Due to the large volume of experiments, sophisticated parameter tuning for ANN and SVR is not feasible.

Most of the software was coded in C++. Two methods, ANN [**?**] and SVR [**?**], were implemented in the *R* programming language but these scripts were invoked from the generic C++ code. All the experiments were performed using a Dell Precision Tower 5820 with an Intel Xeon processor W-2145, 4.5GHz Turbo, 11 Mb cache and 16GB 2666MHz DDR4 memory.

4. Results and Discussion

We conducted a large number of computational experiments. For clarity, we first present some over-arching results, then we discuss separately the aggregated results for the performance of different algorithms, the effect of partitioning the temperature data by discrete intervals, the prediction scores by season and temperature, different fuel use, the size of errors depending on the operational status, commentary on the relative importance of individual features, and finally our observations of the limitations of this approach.

4.1. Results Summary

Looking at a single supermarket using both gas and electricity (SEG), the predicted and real 2017 Summer EDLPs (using 2016 data computed with the kNNR algorithm) are shown in Fig. 2. For this season and year, the best combination of features and number of supermarkets to predict the whole

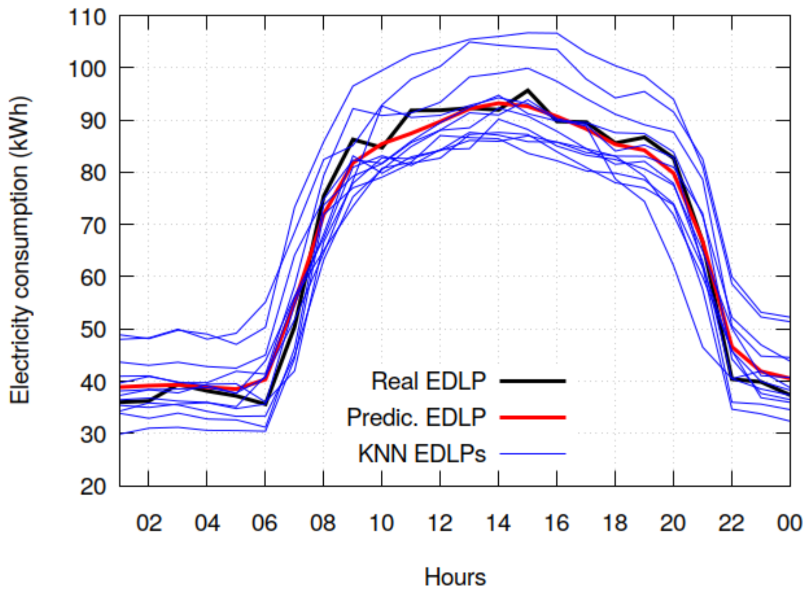


Figure 2: Examples of the EDLPs modelled using kNNR with $k = 12$. The EDLP with the minimum error (the most likely prediction) is shown in red.

of the SEG group are $F = \{\text{GM, Food, Cafeteria}\}$ and $k = 12$ respectively. The blue curves in Fig. 2 are the EDLPs of the k most similar supermarkets based on F, the black and the red curves are the real and predicted EDLP, respectively. The errors for this prediction are ED=14.0 kWh and NP=3.6%. This is the predicted EDLP with lowest ED for all the SEG supermarkets when predicting 2017 Summer EDLPs with the kNNR. The ED (kWh) and NP (%) for all of the SEG group and algorithm are shown in Fig. 3.

The variability between supermarkets is displayed in Fig. 3, with the leftmost being the supermarket with the lowest ED. The median (the 50% position) represents the typical prediction, that being supermarkets with a ED of 33.5 kWh. Fig. 4 shows the real and predicted EDLPs for the best and median-error prediction. In the case of the median-error prediction (Fig. 4b) the predicted EDLPs is an underestimation of the real EDLP. There is only a weak relationship between NP and ED as there are supermarkets sorted by ED and not sorted by NP. The average ED and NP for all 126 SEG supermarkets is $\overline{ED}$=43.5 kWh and $\overline{NP}$ 13.0 %, summarizing the prediction performance over the SEG group.

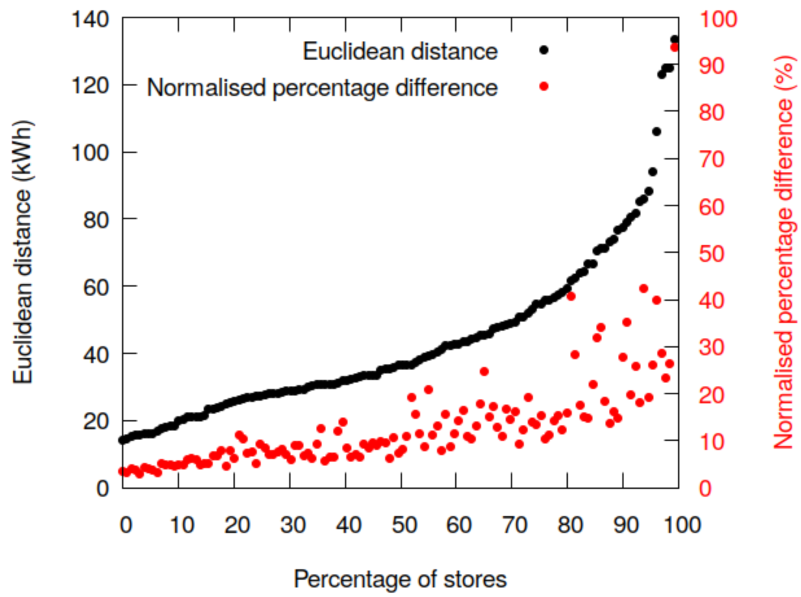
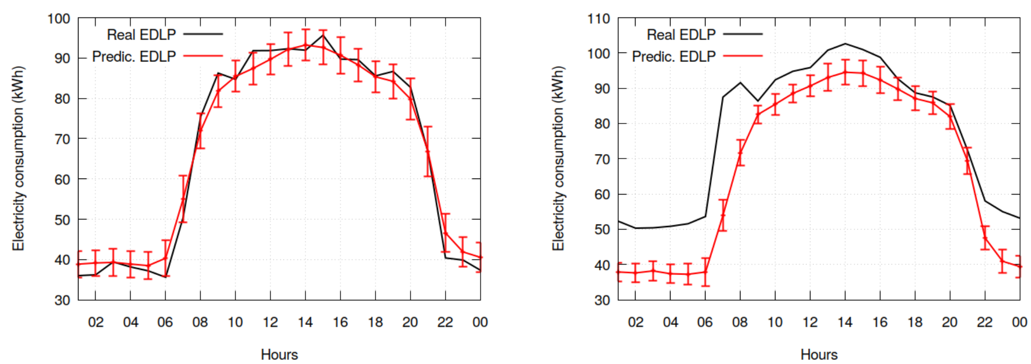


Figure 3: The ED and NP when predicting all of the Summer 2017 EDLPs of the SEG group using 2016 data with the kNNR algorithm. The supermarkets are sorted by ED.



(a) Supermarket with lowest ED

(b) Supermarket with median ED

Figure 4: Prediction of the Summer 2017 EDLPs with lowest and the median ED when predicting all of the SEG group using 2016 data with the kNNR algorithm.

4.2. Algorithm performance and the effect of training data

Considering the range of prediction algorithms, differences among the evaluator scores are not significant for most experiments (Fig. 5, Table A1 and Table A2). For instance, comparing the prediction of Summer 2017 SEG profiles the $\overline{ED}$ score varies from 41.0 kWh obtained with OLS to 45.8 kWh obtained with kNNR. The best results are not always obtained with the same method, but OLS, kNNR and SVR usually obtain lowest errors. Usually, the OLS algorithm obtains the best scores when predicting profiles separated by season, whilst the kNNR method is the best predictor when computing profiles separated by temperature.

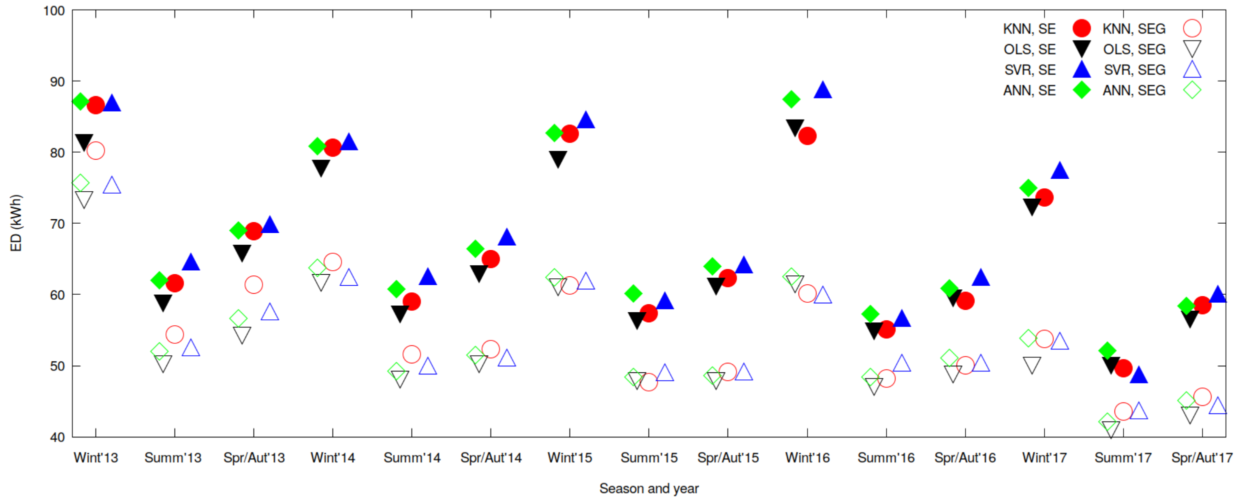
The good performance of the kNNR algorithm compared with more complex algorithms is notable which may be due to the modest size of the data-set. This partially supports the basis of the kNNR *i.e.* similar supermarkets consume energy in similar way. The more complex ML algorithms scale better and may perform better with very large data-sets. On the other hand, the kNNR method is fast and can be used to search larger parameter spaces (k, F).

Table 3 shows the results for SE and SEG using the kNNR algorithm for predicting Summer EDLPs, ncluding the experiments computing the EDLPs of the training set with different numbers of historical years (number of supermarkets are in Table 2). From Table 3, we can see that the best prediction of each year (bold values) is usually obtained using just the previous year as historical data. There are a few exceptions such as for the 2014 SEG group which show that using 2012-2013 profiles for training results are slightly better than using just 2013 data alone. All the supermarkets of Fig. 3 are used to compute the evaluators of the cell located in the first row and column of the SEG sub-table in Table 3.

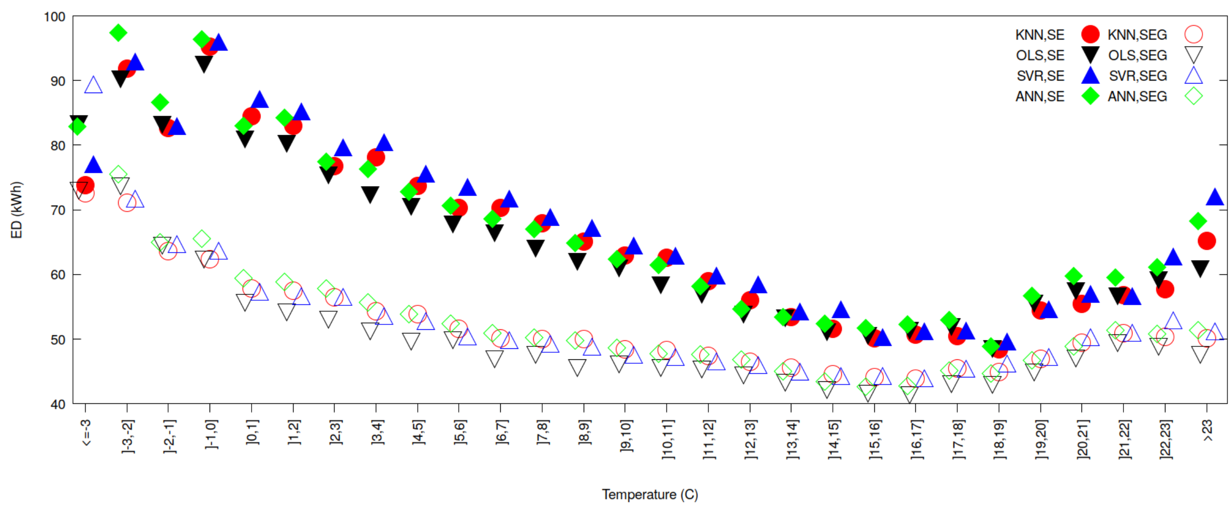
For each method, season and predicted year, the best results obtained with the best combination of historical years (p) are selected (3). The $\overline{ED}$ for all the methods, seasons and years are shown in Fig. 5, where Fig. 5a with Fig. 5b showing the scores for seasonal and temperature experiments, respectively. Table A1 and Table A2 display the MSE for seasonal and temperature experiments, respectively. In comparing the seasonal results for different years, the error usually decreases when predicting EDLPs of more recent years (Fig. 5a). The reason is that the error scales with consumption that decreases with the time. The relative error $\overline{NP}$ also decreased. We suggest that this indicates that the company has sought to harmonise installed equipment in recent years.

Alg	Year	Previous years in Training set, ($\overline{ED}$ (kWh) and $\overline{NP}$ (%)) One	Two	Three	Four	Five
SE	2017	**49.6**/17.0	50.0/**16.9**	51.9/18.1	53.7/20.0	55.0/19.7
	2016	**55.1/18.0**	57.8/19.6	59.8/20.3	60.7/21.1	-
	2015	**57.4/19.9**	59.0/20.7	59.4/20.9	-	-
	2014	**59.0/18.9**	59.2/19.4	-	-	-
	2013	**61.6/19.2**	-	-	-	-
SEG	2017	**43.5/13.0**	44.0/13.0	44.9/13.4	46.3/13.7	46.9/13.9
	2016	**48.2/13.3**	49.6/13.9	51.6/14.6	52.2/14.9	-
	2015	**47.6/14.6**	49.7/15.4	49.8/15.6	-	-
	2014	53.3/15.3	**51.6/14.8**	-	-	-
	2013	**54.4/14.4**	-	-	-	-

Table 3: Prediction results for the SE and SEG groups using the kNNR algorithm and the historical years used. The best results for each year are in bold.



(a) Seasonal results.



(b) Temperature results.

Figure 5: $\overline{ED}$ (kWh) computed using the four ML methods for the seasonal and temperature experiments.

4.3. Using discrete temperature intervals

The temperature data needs to be discretised because of the need to group days with similar temperature conditions (Section 3.1). The error varies depending on the temperature interval in which the profile to predict is computed (Fig. 5b). The error value for the intervals with average temperatures lower than -1 °C and higher than 21 °C is due to the lower number of supermarkets in these intervals. For the intervals from 0 °C to 20 °C, in which the distribution of supermarkets is approximately even and accounts for most supermarkets, the error for the SE and SEG groups show similar behaviour. From left to right in Fig. 5b, we can see that the error starts high for cold temperatures, reducing slowly until it reaches a minimum value for the intervals at approximately 17 °C. After that it increases again showing the influence of the HVAC system.

For very cold temperatures, heating systems are used intensively making predictions more complicated as each supermarket has different thermal conditions and perhaps heating system. For hot temperature intervals (more than 19 °C), the cooling system and the refrigeration appliances can produce the same effect, increasing consumption and the error. Although not surprising, the higher the consumption, the greater the number of appliances, and the greater the variability, the more complicated it is to predict the consumption.

4.4. Partitioning the data by temperature and season

Seasonal and temperature experiments show errors of the same order of magnitude. For instance, the minimum error for the SE group by season (Fig. 5a) is obtained when predicting the Summer 2017 profiles ($\overline{ED} = 48.7$ kWh, using SVR). Meanwhile the minimum error for temperature separation (Fig. 5b) is $\overline{ED} = 48.5$ kWh (using kNNR). There is a similar behaviour of the error for both approaches with respect to the temperature variation. Profiles corresponding to the coldest periods (Winter and for intervals > 5 °C) are predicted less well than for warmest periods (Summer and for intervals < 15 °C). However, the effect of hot temperatures (intervals < 19 °C) which give greater prediction errors, cannot be captured with the seasonal approach. External temperature is a crucial factor in the way supermarkets consume energy and we have already commented that the seasonal separation is a proxy of the temperature separation. Therefore, despite of having sometimes a greater error with the temperature-intervals approach, predicting the EDLPs for new supermarkets with this separation is more useful that using a seasonal profile. Using temperature intervals depends on the availability of daily temperature data.

Comparing the scores that were obtained for each season, Summer profiles were predicted best followed by Spring/Autumn and lastly Winter (Fig. 5a), with this pattern constant for all years and independent of SE/SEG (Fig. 5a). The reasons for this behaviour may be related to the electricity consumption of the heating system as it is used less often in Summer. A fact that supports this assumption is that, in these supermarkets where electrical heating is less important (SEG), the difference of the error between Winter profiles and the other profiles are smaller as happens with the SE group. It also explains the higher error when predicting the Spring/Autumn profiles compared with Summer. Analysis of the temperature results supports this hypothesis.

4.5. Does it matter if a supermarket uses gas-fired heating?

Generally, for the same type of experiments, the errors for the SE group are greater than for the SEG group (Fig. 5a and Fig. 5b). For seasonal experiments and using a relative evaluator such as $\overline{NP}$ the prediction of 2017 Summer profiles using OLS are some of the most accurate predictions with $\overline{NP} = 17.9\%$ and $\overline{NP} = 11.9\%$ for the SE and SEG groups respectively. Likewise for the $\overline{NP}$ evaluator computed over temperature experiments. The reason for this is that variations in heating demand are excluded in SEG and only regulated electricity consumption is computed. Furthermore, the SEG group is larger than the SE set (Table 2) which helps improve the ML prediction. It is expected that most supermarkets will become SE because of the drive for the decarbonisation of heating [**?**].

4.6. Comparing peak/off-peak periods

For peak/off-peak use we analyse the errors during operational times (5am to 10pm) and non-operational times (11pm to 4am) by computing evaluators separately over the two time intervals. For example, the errors to predict the Summer 2017 EDLPs (electricity only) using SVR are $\overline{ED}$=44.4 kWh and $\overline{NP}$=17.4% for the operational periods and $\overline{ED}$=16.4 kWh and $\overline{NP}$=19.7% for the non-operational periods. Considering all the seasonal experiments for all the methods, the average errors are $\overline{ED}$=56.4 kWh and $\overline{NP}$=17.1% for the operational times and $\overline{ED}$=20.2 kWh and $\overline{NP}$=22.5% for the non-operational times.

As the consumption during operational times is higher than for non-operational times, noting the unequal number of hours in the intervals, the relative error, $\overline{NP}$ is a better indicator with which to compare errors than the accumulative real error of $\overline{ED}$. Table 4 shows $\overline{NP}$ the values for operational and non-operational periods averaged over all methods and years. The errors for the non-operational periods are always greater than for the operational periods because the proposed parameter search (Equation 1) minimises the ED between the real and predicted EDLP. Therefore, the method selects the prediction with smaller relative errors in hours with greater consumption. As during non-operational times the electricity consumption is lower than during operational times, reduction of relative error of the latter is prioritised over reduction of relative error of the former.

Trying to predict better the operational times is more difficult, but more useful. Energy use in the non-operational periods is easier to predict since there are fewer human behavioural components contributing to the EDLP. We minimise NP instead of ED (Equation 1) if the relative error is the objective.

	SE group		SEG group	
	Operational	Non-Operational	Operational	Non-Operational
Winter	21.9 (0.5)	30.9 (1.1)	16.6 (0.4)	20.2 (0.7)
Summer	18.1 (0.2)	23.4 (0.5)	13.1 (0.2)	17.4 (0.4)
Spring/Aut	18.9 (0.3)	25.3 (0.6)	13.8 (0.2)	17.3 (0.4)

Table 4: Values for $\overline{NP}$ (%) during operational and non-operational times averaged over all the methods and years. Values in brackets are the standard error.

4.7. Are all features equally useful?

From all the possible features used as predictors (Section 3.1) some are selected more often than others during the feature search process (Equation 1) when considering the whole set of prediction experiments. This means that some features are globally more relevant than others in the prediction. To understand this feature-weighting we analyse only the experiments giving the best results for each combination of algorithm, fuel and temperature/season partition (344 different prediction experiments).

The three features most frequently appearing are Cafeteria area (55.5% of the experiments), Food area (48.2%) and Chilled area (39.8%). Only 52% of the supermarket set have a Cafeteria area, however it is the predictor most frequently selected as the increase of consumption is significant. The Food

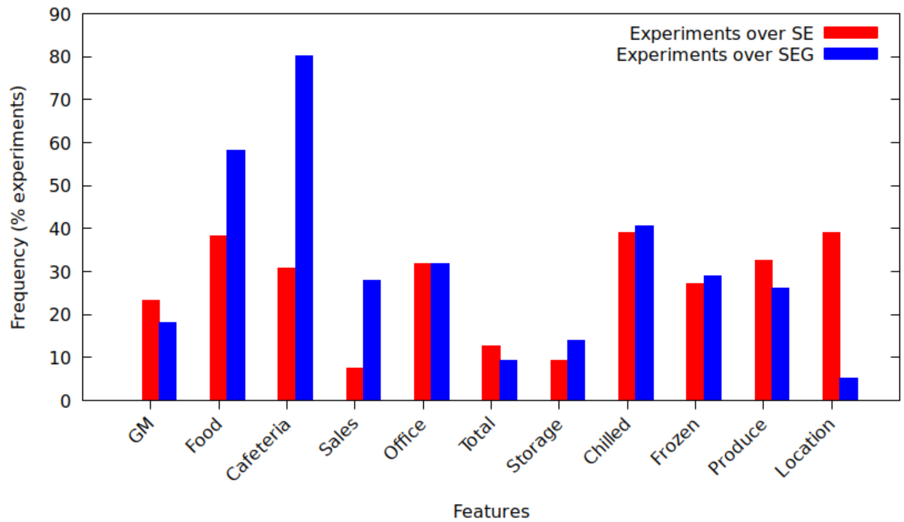


Figure 6: Histogram with the relative frequency of features used to obtain best prediction models for experiments for the SE and SEG groups.

and Chilled areas indicator the number of refrigeration appliances that are responsible of an important part of the electricity consumption.

Interestingly, if we analyse separately the experiments for the SE and SEG groups (177 experiments for each) the frequencies are different for some features. Fig. 6 shows the relative frequency of features used to obtain the best model for all the algorithms and years. The Cafeteria feature appears in 80.2% of experiments for the SEG group, but just 30.8% for the SE group. The Food and Sales area also appear more often in experiments for the SEG group than for the SE group. The Location feature appears in 39.0% of the experiments for the SE group, but in only 5.2% for the SEG group. Most of the experiments to predict consumption for the SE group when daily average temperature was lower than 13 °C has location in the best feature combination. The average number of features used for prediction is 2.9 and 3.4 for SE and SEG, respectively. Seasonal and temperature experiments do not have significant differences in the features frequencies.

4.8. Limitations

Our study has limitations, some understood at the outset, others discovered and quantified during the research. They relate to the use of confidence intervals, and to the nature and availability of data.

The use of confidence intervals helps to model the prediction uncertainty, but there are two limitations in the current implementation. The first is in using techniques that require a large set of training values k, namely OLS, ANN and SVR, to obtain the best results. This yields a large standard deviation. The second limitation is the use of symmetric upper and lower intervals, when different values will be more informative.

The key limitation is the data requirement for ML methods. The errors are not generally very low for all the prediction algorithms and experiments (Fig. 5, Table A1 and Table A2), underlining that the complexity of the problem is related to the data:

Supermarkets vary considerably in total energy consumption. Each is an independent electricity consumer with its own peculiarities *e.g.* location, building features, human factors, and weather conditions, that cannot be completely captured in a model. Moreover, there were no clear criteria to remove any outliers. For instance, the supermarket with the greatest error shown in Fig. 3 (rightmost point) has an unusually large GM area (278 m^2 compared with the average of 48 m^2).

Energy consumption varies over time. Even recent historical data may not be a good guide to future consumption, since changes may arise year–to–year due to weather conditions or refurbishment for example.

The supermarket-set size. The performance of ML algorithms strongly depends on the number of samples (individual supermarkets). The accuracy of our predictions is related to the modest quantity of supermarkets (l< 130 in the separate SE and SEG groups), and not the quantity of the time-series data.

The type of data available. We are limited to what data is available *i.e.* what the supermarket owners are willing to collect or disclose. Accessing more data is desirable *e.g.* the number of customers, technologies used for HVAC and refrigeration, building age, construction type and materials, and insulation levels. However, data collection has a financial cost which must always borne in mind.

Despite this, some individual supermarkets are estimated well. For example, the evaluator scores (Fig. 3) shows there are some supermarkets with low error, with one of the best predictions (Fig. 4a) being a ED of 14.5 kWh and NP of 3.6% (using kNNR). It is possible that the selected features provide good prediction for some supermarkets, but we focused on the feature combination to reduce the average error. When using the OLS method the predicted energy profile was modelled linearly, thus some supermarkets will follow the linear model better than others. Likewise for the more complex SVR and ANN models.

5. Conclusions and future work

We have presented a data-driven method using four ML algorithms to predict the EDLPs of new supermarkets exploiting only historic electricity readings and supermarket features. The data-set comprised six years of hourly electricity readings from 213 UK supermarkets (of one chain), which we partitioned by season and temperature.

The algorithms showed similar prediction scores, where the simplest methods (kNNR and OLS) sometimes out-perform ANN and SVR. In general, the average errors ranged between 12–20% depending on the fuel consumed by supermarkets and season/temperature partition of the readings. However, some EDLPs were accurately predicted (approximately 3% error). We found that warm periods usually were predicted better than cold periods, but the prediction error also increased for very hot intervals (24-hour average $<= 17$ °C). Supermarkets using electricity and gas are better predicted than supermarkets solely using electricity. We suggest that this may be due to the greater variation in the management of HVAC systems when used for heating, compared with using gas.

The features with the strongest effect on the accuracy of the EDLP predictions were the floor areas for Food, Chilled, and Cafeteria. For the SE group the location was also important. As moving to electrical heating is being targeted in the UK [**?**], the relevance of this feature will grow. This can be extrapolated to predict EDLPs for supermarkets in countries with hot climates where the cooling system has greater weight in the electricity consumption than in the UK.

Our work suggests that accuracy increases with the store sample size; ML methods ideally need more samples than the 213 supermarkets used in this study. The main advantage of our ML approach is the simplicity of using a small number of easily obtained parameters (features) to compute useful information (energy use) for the whole portfolio of stores, not just for an individual store. Additionally, as data-driven models the proposed methods can be extended to predict EDLPs of any type of store, not just supermarkets, when data is available.

There are several research lines that follow from this work. Other ML methods can be tested and the confidence intervals can be improved to better model the uncertainty. Furthermore, different temperature intervals can be investigated, *e.g.* 2 °C intervals or an interval width depending on the consumption variation. Finally, there may be merit to developing independent models for the operational and non-operational periods to account for the different behaviour.

Acknowledgements

We are grateful the supermarket company for providing the data used in this work. The contribution by R. Granell and D.C.H Wallom was partially funded by University of Oxford EPSRC Impact Acceleration Account (grant no. EP/R511742/1).

Appendix A

Year	Season	Stores with elec. only KNN	OLS	SVR	ANN	Stores with elec. and gas KNN	OLS	SVR	ANN
2013	Wint	498.3	366.5	522.1	466.7	346.9	293.5	336.2	319.6
	Sum	317.0	243.2	394.2	290.7	174.2	137.5	159.6	149.4
	Spr/Aut	323.2	259.4	354.3	339.1	206.5	154.8	182.6	174.2
2014	Wint	357.3	334.1	371.6	362.7	223.8	217.2	226.1	227.9
	Sum	236.7	202.0	288.4	226.5	142.8	131.8	131.1	129.5
	Spr/Aut	220.6	207.0	266.6	236.2	145.3	137.9	139.0	144.4
2015	Wint	393.8	323.0	406.7	384.9	210.7	212.3	219.0	217.7
	Sum	221.3	194.2	266.6	232.8	121.7	125.9	132.2	126.2
	Spr/Aut	228.2	206.6	260.6	221.7	124.7	122.3	127.7	123.3
2016	Wint	364.3	342.4	413.6	403.9	183.7	191.9	183.5	193.4
	Sum	177.0	171.3	272.3	213.0	112.5	108.6	125.5	114.7
	Spr/Aut	194.4	193.5	219.5	193.7	124.9	117.8	127.1	126.9
2017	Wint	308.4	280.5	366.7	314.1	176.5	160.1	181.5	173.4
	Sum	189.9	139.7	187.1	165.3	103.9	92.6	107.6	96.8
	Spr/Aut	195.8	174.4	224.7	189.9	122.6	108.5	121.7	116.5

Table A1: Prediction results using the MSE (kWh) evaluator for the algorithms over seasonal experiments for all years and store types.

Temp (°C)	Stores with elec. only KNN	OLS	SVR	ANN	Stores with elec. and gas KNN	OLS	SVR	ANN
≤ -3	251.9	367.5	255.2	321.1	300.5	376.8	529.9	575.0
]-3,-2]	392.2	430.0	424.1	472.6	388.6	447.0	344.1	401.7
]-2,-1]	324.2	490.0	329.4	354.9	326.5	368.7	339.7	330.1
]-1,0]	495.7	666.0	502.8	492.5	275.4	341.0	279.2	283.6
]0,1]	423.7	382.2	456.6	356.1	203.6	255.6	208.8	205.8
]1,2]	397.5	452.0	437.6	385.0	205.9	253.9	207.3	212.4
]2,3]	347.4	360.9	379.3	314.2	200.8	227.1	202.9	204.9
]3,4]	408.2	350.6	472.4	319.7	182.2	211.4	183.7	186.4
]4,5]	332.6	326.9	370.9	272.8	173.5	195.6	174.1	171.0
]5,6]	265.9	293.0	310.8	251.0	162.2	195.3	159.9	162.5
]6,7]	281.9	289.8	341.0	258.4	149.9	176.1	153.4	152.5
]7,8]	261.9	273.1	298.2	253.1	145.1	179.5	146.2	144.5
]8,9]	248.2	292.7	294.3	232.2	138.9	184.5	139.2	139.6
]9,10]	212.2	269.1	274.9	203.4	133.0	189.6	130.8	130.5
]10,11]	241.6	269.1	287.5	229.7	128.6	189.0	127.7	128.1
]11,12]	199.7	259.8	234.4	180.3	122.4	196.1	124.9	123.4
]12,13]	195.4	235.9	234.9	174.9	120.4	195.3	122.2	122.8
]13,14]	173.4	230.0	185.0	157.5	114.2	174.8	113.8	112.8
]14,15]	176.0	213.8	210.3	164.4	112.7	154.1	112.8	107.8
]15,16]	170.3	166.4	185.9	162.1	108.9	130.1	108.2	100.5
]16,17]	168.0	190.3	186.6	169.8	106.8	127.1	106.6	97.7
]17,18]	184.2	194.1	193.1	178.7	114.0	130.0	117.6	112.9
]18,19]	147.2	171.4	207.2	165.3	105.7	115.8	111.0	107.2
]19,20]	188.4	221.9	244.2	217.4	121.4	120.0	127.3	121.8
]20,21]	194.9	279.7	278.5	253.3	129.8	154.3	134.6	127.6
]21,22]	231.7	246.3	263.4	235.3	132.6	186.5	132.9	133.4
]22,23]	247.8	273.5	303.7	257.1	135.6	174.4	152.5	135.2
>23	327.7	246.6	408.9	324.3	141.9	163.4	146.5	152.3

Table A2: Prediction results using the MSE (kWh) evaluator for the algorithms over temperature experiments.