Unsupervised machine learning techniques for energy consumption tariff design

Tsakiris Georgios
Aristotle University of Thessaloniki
Thessaloniki, Greece
tsakiris@csd.auth.gr

Virtsionis-Gkalinikis Nikolaos
Aristotle University of Thessaloniki
Thessaloniki, Greece
nvirtsion@csd.auth.gr

Mischos Stavros
Aristotle University of Thessaloniki
Thessaloniki, Greece
mischossg@csd.auth.gr

Vrakas Dimitrios
Aristotle University of Thessaloniki
Thessaloniki, Greece
dvrakas@csd.auth.gr

ABSTRACT

Dealing with electricity demand fluctuations throughout peak and off-peak periods is challenging for electricity companies. During peak demand times, the grid should be able to match the high consumer needs. Conversely, minimal usage during off-peak periods leads to underutilization of generation capacity. This imbalance challenges utilities to ensure sufficient capacity and devise fair pricing models. The Time-of-Use (ToU) pricing model has emerged as a viable solution in many countries, encouraging consumers to shift their energy consumption from expensive peak hours to more affordable off-peak periods. To this end, this paper proposes unsupervised machine learning methods for designing ToU tariffs using only energy consumption time series data. Additionally, a new metric is introduced to evaluate the adaptability of the ToU methods to fluctuations in energy consumption. To validate the implemented techniques, public datasets from different countries were used.

KEYWORDS

Time-Of-Use pricing, residential energy consumption, fuzzy clustering, unsupervised machine learning

ACM Reference Format:

Tsakiris Georgios, Virtsionis-Gkalinikis Nikolaos, Mischos Stavros, and Vrakas Dimitrios. 2018. Unsupervised machine learning techniques for energy consumption tariff design. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 10 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Electricity pricing is influenced by various factors such as demand fluctuations, availability of generation sources, and operational costs. Naturally, high energy demand leads to increased energy prices due to the incorporation of more expensive generation sources to meet the increased load. Energy retailers can motivate customers to adjust their consumption patterns through Demand Response (DR) programs [31], which involve modifying electricity usage in response to price signals or incentives. ToU pricing, a form of DR, encourages load shifting by imposing higher prices during peak hours, prompting consumers to redistribute their energy usage to lower-cost periods [5, 11, 15].

The use of ToU tariffs could be beneficial for both consumers and energy companies [19]. Firstly, consumers would experience lower electricity costs. Additionally, by avoiding periods of excessively high demand, overall energy production and management costs could be reduced, leading to lower wholesale electricity prices. Moreover, ToU tariffs could improve the economic viability of various distributed energy sources, such as solar panels, energy storage and electric vehicles. For instance, household owners that produce electricity via solar panels might receive compensation for each kWh of energy they generate.

Most state-of-the-art techniques produce ToU tariffs based on peak and off-peak pricing structures. Their goal is to incentivize consumers into shifting their energy demand, decrease power grid's pressure during production and distribution, and ensure its stability. However in the literature, there was a lack in ways to evaluate the quality of a developed ToU tariff schema. As far as knowledge is concerned, there is no available metric that can estimate how well a ToU tariff is aligned with the consumption patterns of the time series. Furthermore, the majority of ToU tariffs production algorithms were usually tested only on a single dataset to prove its basic functionality.

In this paper, new unsupervised machine learning algorithms for generating ToU tariffs are introduced, based on forecasted data of the next day. These new algorithms are applied in various public datasets for validation. These datasets originate from different countries, which offers a more reliable statistical analysis on the results. This work also proposes a new metric for the evaluation of the adaptability of ToU tariffs creation methods to various fluctuations of the energy consumption time series. Tariffs, must always be appropriately aligned with the valleys (low consumption periods), the flat periods (medium consumption periods) and the spikes (high consumption periods) of the consumption time series. The proposed metric provides a way to estimate how well ToU tariffs are adjusted in these consumption periods and also how fast or

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

slow the adjustment happened. Using this metric, researchers can compare different ToU tariff-design algorithms, which can lead into more accurate results.

2 RELATED WORK

In the pursuit of optimizing dynamic tariff design for energy consumption, various studies have investigated methodologies employing fuzzy logic and clustering techniques. This section provides an overview of the literature, highlighting key findings and methodologies utilized in previous research.

Dynamic tariff schemes have been explored extensively, with researchers emphasizing the importance of variable energy charges throughout the day to manage load levels effectively. Studies by Shrinivasan et al. [29] and Soland et al. [30] have demonstrated the development of dynamic tariffs aimed at encouraging consumers to adjust their energy usage patterns. These tariffs typically involve higher energy charges during peak periods and lower charges during off-peak periods, incentivizing consumers to shift their consumption habits [18].

Optimal ToU tariff design under uncertainty has also garnered attention in the literature. Previous research efforts have proposed stochastic programming models to manage price changes and encourage load shifting among customers. Game-theoretic approaches have been employed to optimize ToU pricing strategies and achieve a Nash equilibrium between companies and consumers. Additionally, the integration of ToU pricing with economic dispatch problems has been explored to enhance network reliability and customer benefits [4, 7, 12, 14, 35].

Clustering techniques have been utilized to address challenges associated with time period segmentation for tariff design. Studies have implemented various clustering methods, such as linear integer programming and Gaussian mixture models, to partition time periods and design ToU tariffs for domestic customers [22, 27].

Moreover, the literature underscores the significance of methods based on fuzzy membership functions in peak-valley time division models. These methods utilize fuzzy c-means (FCM) clustering algorithm based on membership functions to address time period partition challenges associated with ToU tariffs [9, 21, 34, 36].

However, the clustering effect is sensitive to initialization, prompting research into improving clustering algorithms. Various methods have been proposed, including fuzzy c-means with discriminative embedding to address suboptimal results [24]. Modifications to the initialization process, such as setting small non-negative values and adjusting parameters, have been explored to stabilize clustering initialization [13, 26].

In conclusion, previous research efforts have laid the groundwork for dynamic tariff design using fuzzy logic and clustering techniques. While promising, further research is needed to integrate these methodologies into real-world energy market scenarios and assess their practical feasibility and effectiveness [10, 33].

3 CONTRIBUTIONS

The proposed work contributes to the field of ToU tariff design in the following points. Firstly, with a proposed algorithm for tariff design. The algorithm, based on the concepts of fuzzy systems and clustering, produces the corresponding tariff values using only the daily consumption time series as input. This underscores the advantage of unsupervised learning techniques, as they facilitate the identification of dynamic pricing patterns at a reduced computational expense. Next, with the investigation of unsupervised machine learning algorithms versus custom-tailored heuristics methods for the problem of tariff design. Finally, a novel performance evaluation metric is introduced to measure how fast the algorithm adapts to relevant fluctuations of the load consumption.

The rest of the paper is organized as follows. Section 4 contains a concise description of the datasets used to validate the proposed techniques. The methodology that was followed is thoroughly described in Sections 5 and 6. Section 7 presents the experiments that were conducted. Section 8 presents the results produced after qualitative and quantitative evaluation. Section 9 contains the relevant conclusions and ideas for future work.

4 DATA

This section presents the data utilized in designing and comparing dynamic tariffs for electricity consumption. The data primarily consists of smart meter readings obtained either at the individual consumer level or aggregated at the network level. It is important to note that due to the limited use of dynamic pricing during data mining, the actual pricing in the datasets is unknown. Table 1 provides a comprehensive overview of the datasets utilized in this study.

4.1 Low Carbon London Dataset

The Low Carbon London dataset [28], was created by UK Power Network, spans from November 2011 to February 2014 and encompasses residential consumers in London. It comprises electricity consumption data from 5567 residential consumers, recorded at 30-minutes granularity. The consumers were categorized into two subgroups based on their tariff structures: 4334 consumers under fixed tariffs and 1199 consumers under dynamic tariffs. Additionally, the dataset includes tariff details for the respective subgroups, alongside supplementary demographic, weather, and calendar information. Notably, only electricity consumption data at the network level was utilized in this analysis.

4.2 CAMSL Dataset

The CAMSL dataset [20], was created collaboratively by Looop Inc. and SMAP ENERGY Limited, aimed to introduce dynamic pricing in Tokyo, Japan. The project spanned from July 1, 2017, to December 31, 2018, involving 1423 residential consumers, with data recorded at a 30-min granularity. Consumers were segmented into three distinct tariff subgroups: 400 consumers under controlled pricing, 1023 consumers under dynamic pricing, and 3337 consumers under fixed pricing. Similar to the Low Carbon London dataset, the CAMSL dataset includes tariff specifications for the subgroups, along with supplementary demographic, weather, and calendar data. For this study, only electricity consumption data at the network level was utilized.

4.3 Open Power System Data

The Open Power System Data platform [25] provides open-access data crucial for power system design and modeling. It encompasses

Table 1: Overview of Datasets

Dataset Name	Duration	Geography	Consumer Count	Sampling Frequency
Low Carbon London Dataset	Nov 2011 - Feb 2014	London, UK	5567	30 min
CAMSL Dataset	Jul 2017 - Dec 2018	Tokyo, Japan	1423	30 min
Open Power System Data	Jan 2015 - Sep 2020	Great Britain, Ireland, Austria	N/A	15, 30, 60 min

data on total electricity consumption, production, and storage of solar and wind energy, as well as tariff information for 32 European countries and select neighboring nations. These datasets offer various sampling periods, including 15 minutes, 30 minutes, and 1 hour. For this paper, data from Great Britain, Ireland, and Austria were extracted, covering the period from January 2, 2015, to September 30, 2020, with a sampling frequency of 1 hour.

5 METHODOLOGY

The imperative need to increase the utilization of renewable energy sources and mitigate energy footprints has spurred extensive research efforts toward sustainable solutions [3]. Among these, demand response through dynamic techniques has emerged as a promising avenue [1]. Historically, diverse methodologies have been proposed to design dynamic pricing, often relying on statistical models [8]. However, these approaches typically demand expertise from skilled scientists and engineers. Leveraging artificial intelligence (AI) and machine learning (ML) techniques for dynamic pricing design offers a viable alternative, as it can harness insights solely from historical data, circumventing the necessity for specialized expertise [10, 33]

Below, some of the principal techniques employed in dynamic pricing design are outlined:

5.1 Heuristic Thresholding

Heuristic Thresholding operates on classical principles of dynamic pricing design, leveraging statistical analysis and threshold setting to define distinct pricing tiers [8]. In this approach, thresholds are established following an analysis of historical data, effectively partitioning pricing levels.

To tailor this technique to the Low Carbon London dataset, an analysis was conducted to determine optimal threshold values for tariffs. This analysis involved normalizing mean daily consumption by dividing it by the number of consumers enrolled in the scheme and the mean price of consumption. Notably, normalization facilitated a direct comparison between the two consumption curves.

Figure 1 illustrates both mean consumption and total normalized consumption, revealing a consistent percentage difference between the two curves. Upon dataset examination, it was observed that the maximum mean electricity consumption corresponds to approximately 80% of total normalized consumption. This observation led to the decision to designate this value as the benchmark for establishing the upper tariff limit. Conversely, the minimum price served as the reference point for determining the lower tariff limit. Equations 1 and 2 are utilized to compute the thresholds of the tariffs.

$$T_L = (1 + low_{coef}) \times min_{dc} \tag{1}$$

$$T_U = (1 + high_{coef}) \times max_{dc} \times 80\% \tag{2}$$

Where:

- T_L is the lower tariff limit.
- T_U is the upper tariff limit.
- low_{coef} and $high_{coef}$ represent the percentage difference from the reference points.
- min_{dc} represents the minimum daily consumption
- max_{dc} represents the maximum daily consumption

5.2 K-Means

The k-means algorithm [17] is an iterative clustering method employed to partition a dataset into 'k' distinct clusters based on similarity measures. It proceeds iteratively by initially assigning data points to the nearest cluster centroid and recalculating centroids based on the mean of the data points assigned to each cluster. The algorithm commences by randomly initializing 'k' centroids within the feature space. In each iteration, it assigns each data point x_i to the cluster with the nearest centroid c_j based on a distance metric, typically the Euclidean distance formula [2]:

$$d\left(x_i, c_j\right) = \sqrt{\sum_{m=1}^{n} \left(x_{im} - c_{jm}\right)^2} \tag{3}$$

Here, 'n' signifies the number of dimensions in the feature space, x_{im} denotes the m-th component of the i-th data point, and c_{jm} signifies the m-th component of the j-th centroid. Following the assignment of all data points, centroids are updated by computing the mean of all points assigned to each cluster. This iterative process continues until convergence criteria are met, such as a minimal change in centroid positions or reaching a maximum number of iterations. The outcome comprises 'k' clusters, each represented by a centroid, with data points assigned to the cluster whose centroid they are closest to.

Below is the pseudocode for the k-means algorithm:

5.3 Fuzzy C-Means

The Fuzzy C-Means (FCM) [32] algorithm emerges as a prominent method within cluster analysis, distinguishing itself from traditional clustering algorithms by assigning membership values to data points, indicating the degree of association with each cluster. Mathematically, FCM endeavors to minimize the objective function J, which quantifies the overall fuzziness of the clustering:

$$J = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} \cdot \left\| x_i - v_j \right\|^2 \tag{4}$$

Where:

- n denotes the number of data points,

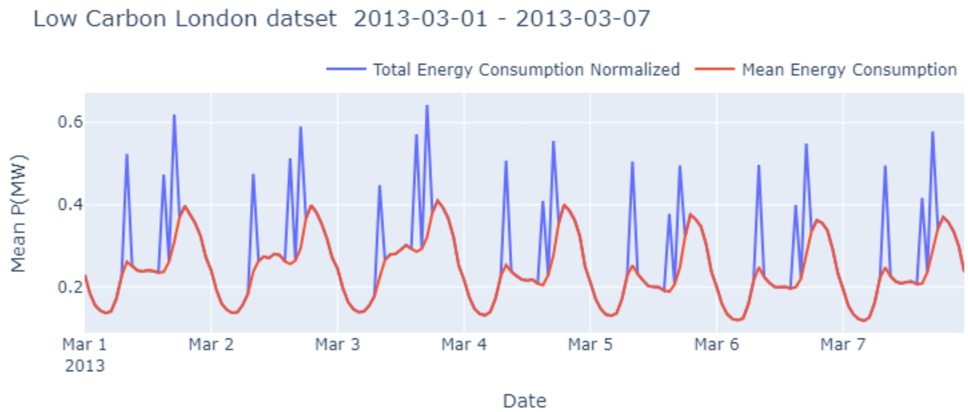

Figure 1: Mean and Total Normalized Energy Consumption in Low Carbon London dataset

Algorithm 1 K-Means

```
Initialize centroids randomly within the feature space.
while not Convergence criteria are met do
    for each data point do
        Assign the point to the nearest centroid based on the distance metric.
    end for
    for each centroid do
        Update the centroid by computing the mean of all points assigned to it.
    end for
end while
return Final centroids and cluster assignments.
```

- c represents the number of clusters,
- u_{ij} signifies the membership of data point x_i in cluster j,
- v_j denotes the centroid of cluster j,
- m is a weighting exponent typically set to 2 for crisp partitioning, yet adjustable for fuzzier partitions.

Initially, FCM assigns membership values randomly to each data point. It then proceeds to iteratively update cluster centroids and membership values alternately until convergence. The updated equations are delineated as follows:

Membership Update:

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - v_j\|}{\|x_i - v_k\|} \right)^{\frac{2}{m-1}}} \tag{5}$$

Centroid Update:

$$v_j = \frac{\sum_{i=1}^{n} u_{ij}^m \cdot x_i}{\sum_{i=1}^{n} u_{ij}^m} \tag{6}$$

Throughout each iteration, data points are re-evaluated for membership values based on their proximity to cluster centroids, with closer points accorded higher memberships. Concurrently, centroids are recalibrated through weighted averages of data points, where the weights are determined by membership values. This iterative process persists until membership values stabilize, indicating well-defined clusters.

FCM exhibits versatility and efficacy in handling intricate datasets featuring overlapping clusters or noisy data. By permitting soft assignments, it adeptly captures the inherent uncertainty prevalent in real-world data. Nonetheless, FCM's performance is contingent upon initializations and parameter selections, necessitating meticulous tuning for optimal outcomes. Despite its limitations, FCM maintains widespread utilization across diverse domains such as pattern recognition [6], image segmentation [16], and data mining [23], owing to its capability to furnish meaningful clusterings across varied contexts.

Below is a pseudocode representation of the Fuzzy C-Means algorithm:

Algorithm 2 Fuzzy C-Means (FCM)

```
Initialize: Choose the number of clusters (c), weighting exponent (m), and terminate threshold (ϵ)
Randomly initialize cluster centroids (v_j)
while not Convergence do
    for each data point x_i do
        for each cluster centroid v_j do
            Compute membership value u_ij using the membership update equation
        end for
    end for
    for each cluster centroid v_j do
        Update centroid coordinates using the centroid update equation
    end for
    Compute the change in cluster centroids
    if change in centroids is less than ϵ then
        Convergence = True
    end if
end while
```

Figure 2: Energy Consumption, Membership Function, and FCM Tariffs in Low Carbon London dataset

6 IMPROVING DYNAMIC PRICING STABILITY OF FUZZY C-MEANS

This subsection outlines the corrections proposed to enhance the stability and reliability of FCM in the context of dynamic pricing. In particular, Correction 1 addresses abnormal periods in clustering significance, while Correction 2 focuses on managing multiple pricing changes within short time intervals.

From the analysis depicted in Figure 2, which showcases the Low Carbon London dataset for 05/03/2012, it becomes evident that the FCM algorithm encounters challenges in providing high accuracy answers during specific time slots, namely 07:00, 9:00, 18:00, and 22:00. Moreover, the algorithm's pricing exhibits notable fluctuations within the 15:00 to 21:00 interval.

6.1 Correction 1: Addressing Abnormal Periods

To mitigate the impact of uncertainty in clustering significance during abnormal periods, a threshold value is computed for each time point based on the variability in the membership function. Specifically, the marginal value threshold is calculated as the product of the difference between the maximum and minimum membership function values and a coefficient representing the percentage of the boundary between these extremes (Equation 7). Abnormal periods are then identified where the difference between the two largest membership function values falls below the computed threshold (Equation 8). The maximum threshold value within these abnormal periods is set to mitigate the impact of uncertain clustering on dynamic pricing stability.

$$T_h = (max_h - min_h) \times \text{coef}, s.t.h \in [0, 23] \tag{7}$$

Where:

- T_h is the computed threshold between maximum and minimum values of the membership function at time h.
- max_h is the maximum value of the membership function at time h.
- min_h is the minimum value of the membership function at time h.
- coef is a coefficient denoting the percentage of the boundary between the maximum and minimum value

$$\begin{gathered}\text{Current Period} = \text{Abnormal Period} \\ \text{if } \text{diff}_h < T_h\end{gathered} \tag{8}$$

6.2 Correction 2: Managing Multiple Pricing Changes

This correction addresses the challenge of multiple pricing changes occurring within short time intervals. Initially, time points exhibiting unit changes in pricing are identified. Subsequently, instances of multiple pricing variations between consecutive time points are determined. To manage these variations, an adjustment is made to intermediate prices based on the prevailing tariff values at neighboring time points. Specifically, intermediate prices are aligned with the prevailing tariff values, ensuring smoother transitions and reducing pricing fluctuations (Equation 9).

$$\begin{array}{ll} p_{h-1} = p_h & \text{if } p_{h-2} = p_h \\ p_{h+1} = p_h & \text{if } p_h = p_{h+2} \\ h \in [0, 23], & h \in \mathbb{Z} \end{array} \tag{9}$$

Where:

- h is the hour of the day
- p_h is the tariff price at time h.

Figure 3 provides a visual representation of the aggregate consumption values alongside the dynamic tariffs of Fuzzy C-Means and Corrections 1, and 2 within the Low Carbon London dataset. The incorporation of these corrections enhances the reliability and stability of Fuzzy C-Means in dynamic pricing applications.

7 EXPERIMENTS

In this section, we detail the experiments conducted to assess the effectiveness of the designed methods discussed in Sections 5 and 6 using the Low Carbon London, CAMSL, and datasets from Great Britain, Ireland, and Austria extracted from the Open Data Platform.

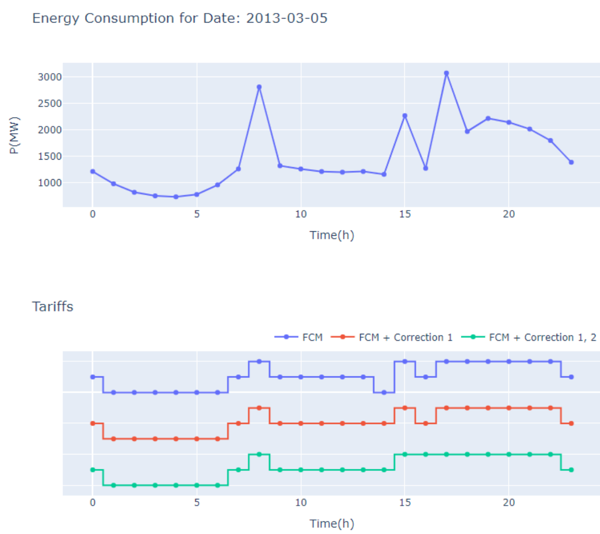

Figure 3: Dynamic Tariffs in Low Carbon London Dataset with Fuzzy C-Means Corrections

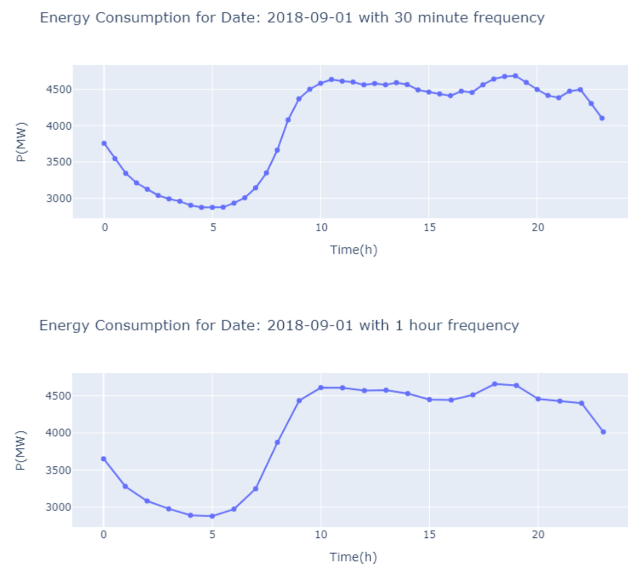

Figure 4: Energy Consumption in Ireland for 30-min and 1-hour

7.1 Preprocessing

The preprocessing step is necessary to bring data in the appropriate form to design dynamic prices and make predictions. For the datasets mentioned, missing values were filled using hourly profiles, and aggregation of consumer consumption was performed for the Low Carbon London and CAMSL datasets to calculate the total consumption of the network.

After analyzing the datasets, Figures 4 and 5 showcase the consumption of Ireland and Austria for different sampling frequencies. It's noted that there are no significant changes in the shape of the waveform. Therefore, it was decided to use a 1-hour granularity to better understand the pricing resulting from the algorithms. This decision was made due to the insignificant changes observed in the consumption time series at 15-minute, 30-minute, and 1-hour intervals.

7.2 Parameter Settings

To create dynamic pricing, algorithms from Sections 5 and 6 were employed. For Heuristic Thresholding, coefficients low_{coef} and $high_{coef}$ were determined through statistical analysis, resulting in optimal pricing with $low_{coef} = 1.45$ and $high_{coef} = 0.90$. For k-Means, no specific hyperparameter tuning was performed as the results were consistent across different settings. However, for Fuzzy c-Means, a grid search was conducted to identify suitable hyperparameters, resulting in the following hyperparameters:

$$m = 7.0, error = 0.0005, max_{iter} = 1000 \tag{10}$$

Additionally, corrections were implemented to optimize the uncertain Fuzzy c-Means tariffs and prevent rapid consumption changes. The coefficients for these corrections were determined experimentally, with $coef = 0.35$ yielding optimal uncertainty improvement for Correction 1. Correction 2, which smoothes changes between low and high tariffs, did not require parameterization.

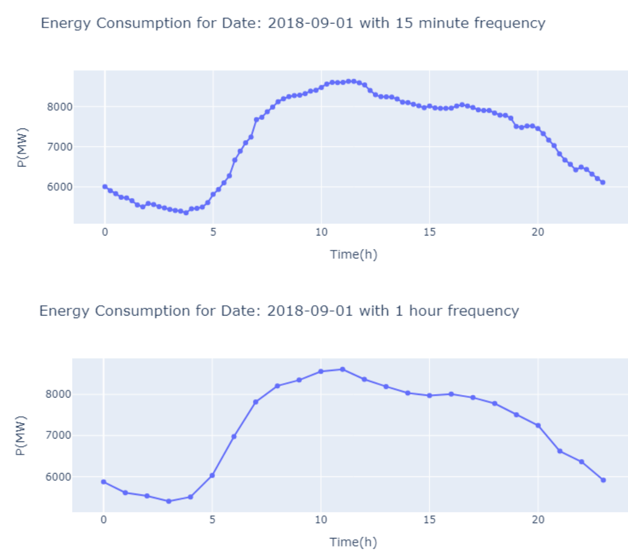

Figure 5: Energy Consumption in Austria for 15-min and 1-hour

7.3 Experiment Setup

The experiments were conducted using the extracted datasets, and the unsupervised machine learning algorithms were implemented with the determined parameter settings. The experiments aimed to observe the performance of each algorithm in identifying periods of high and low demand and adjust tariffs accordingly.

Furthermore, in devising the tariffs, the proposed algorithms leveraged daily data inputs. This highlights the advantage of unsupervised learning techniques, as they enable the identification of dynamic pricing patterns at a reduced computational expense.

The aforementioned datasets served as the basis for the experimental procedure. Heuristic Thresholding was applied as a statistical method for the design of dynamic pricing, while K-Means and Fuzzy C-Means algorithms were utilized as Unsupervised Machine Learning methods. Additionally, Corrections 1 and 2 were applied as a heuristic approach to design dynamic tariffs more favorably from the provider's point of view and to avoid destabilization of the network.

Moreover, to enhance the adaptability of the system, three levels of pricing were created to update the Demand Response elasticity. This decision aimed to provide a more nuanced approach to tariff adjustments based on varying levels of demand.

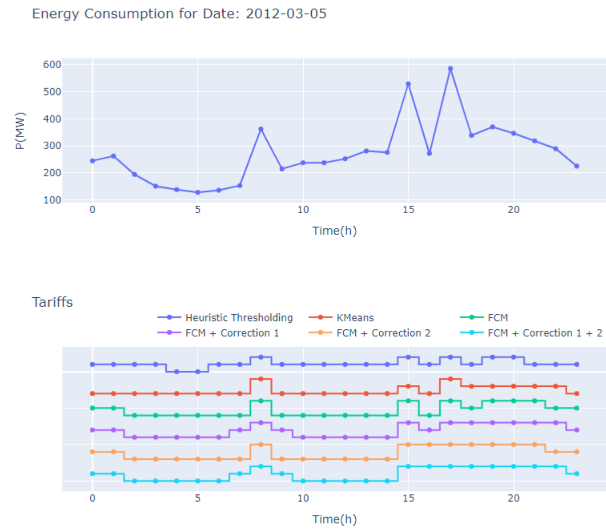

Figure 6: Energy Consumption and Tariffs in Low Carbon London Dataset on 05/03/2012

8 RESULTS

8.1 Qualitative evaluation

Figure 6 illustrates consumption and dynamic tariffs for the Low Carbon London dataset on 05/03/2012, demonstrating that all algorithms effectively identify periods of increased demand (High Tariffs). However, Heuristic Thresholding displays more Normal Tariffs and struggles to identify periods of low tariffs. K-Means identifies high-demand periods but exhibits tariff inaccuracies, as observed at 15:00. On the other hand, Fuzzy c-Means, identifies effectively periods of increased demand, while corrections create smoother transitions between low and high tariffs.

Additionally, it becomes clear that Fuzzy C-Means offers tariffs that do not favor the provider, as they allow the consumer to shift consumption from 15:00 and 17:00 to 16:00 to reduce their bill. This shift poses numerous risks, including network destabilization and potential blackout. Therefore, Fuzzy C-Means with Corrections 1 and 2 offer ideal tariffs for the provider and help to stabilize the network.

Figure 7 showcases consumption and dynamic tariffs for the CAMSL dataset on 20/05/2018, with all techniques successfully identifying high and low demand periods. Fuzzy c-Means, along with corrections, demonstrate robust performance in identifying demand patterns and adjusting tariffs accordingly.

Figures 8, 9, and 10 depict consumption and dynamic tariffs for Great Britain, Ireland, and Austria on 01/09/2018, respectively, demonstrating consistent performance across all techniques in identifying periods of high and low demand. Fuzzy c-Means, supplemented by corrections, exhibit reliable performance in adapting tariffs to varying demand levels

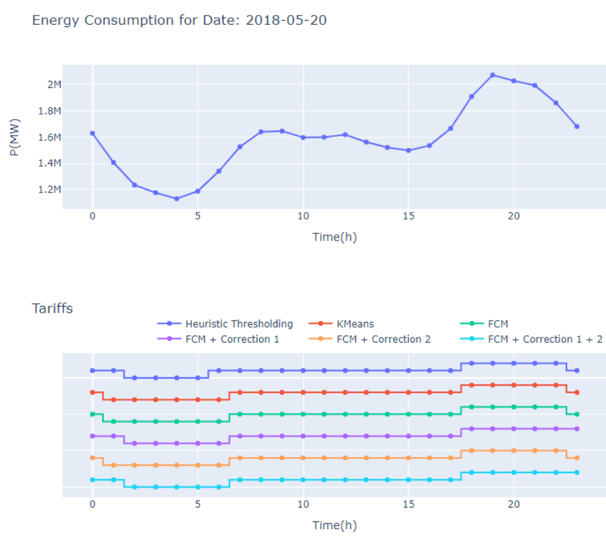

Figure 7: Energy Consumption and Tariffs in CAMSL Dataset on 20/05/2018

8.2 Quantitative Evaluation

8.2.1 Evaluation Metric Design.

To evaluate the efficacy of the tariff design algorithms, a new metric was created. The aim was to ascertain the precision of the algorithms in adapting tariffs during periods of heightened ("peak") and reduced ("valley") energy consumption.

Illustrated in Figure 6 are energy consumption patterns for 05/03/2012, drawn from the Low Carbon London dataset. Through

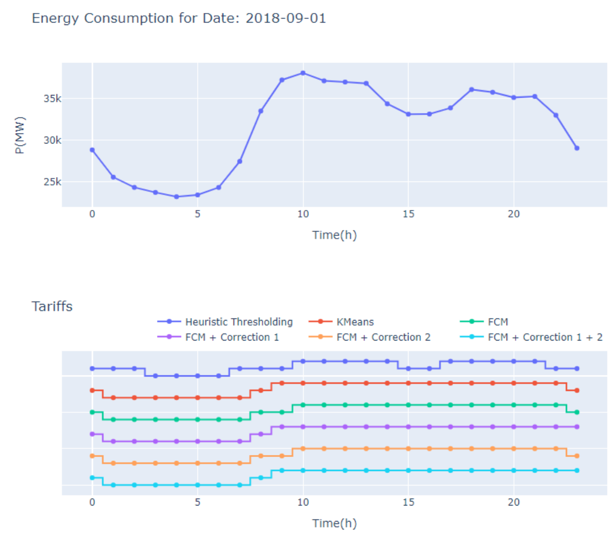

Figure 8: Energy Consumption and Tariffs in Great Britain on 01/09/2018

Figure 10: Energy Consumption and Tariffs in Austria on 01/09/2018

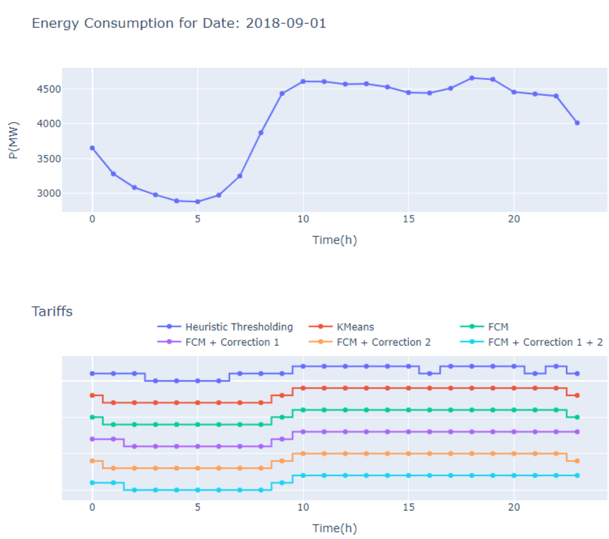

Figure 9: Energy Consumption and Tariffs in Ireland on 01/09/2018

a detailed examination of this data, we identified valley periods between 02:00 - 07:00 and 10:00 - 13:00, alongside peak periods at 08:00, 15:00, and 17:00.

The novel metric emphasizes two crucial aspects: the alignment between peak periods and the consistency of tariff adjustments, denoted by penalties. Initially, consumption data was segmented into distinct periods by detecting local peaks and valleys, thereby defining "valleys" and "peaks".

Segmenting the consumption data in Figure 6 yielded intervals: 00:00 - 01:00, 01:00 - 08:00, 08:00 - 15:00, 15:00 - 17:00, and 17:00 - 23:00. While local minima were observed at 14:00, 16:00, and 18:00, the first and last were considered insignificant due to minimal variation. Notably, the value for 16:00 was encompassed within the 15:00 - 17:00 period, regarded as an abnormal period.

Subsequently, we determined the dominant pricing category for each segment and identified the lag of tariff adjustments from the segment's boundaries. Figure 6 depicts the dominant pricing category for each segment, accompanied by corresponding delays in tariff adjustments from segment boundaries.

Following this, Equation 11 presents the proposed metric:

$$\text{metric} = \text{overlap} - p_1 \times d_s - p_2 \times d_e \tag{11}$$

Where:

- *overlap*: denotes the alignment of dominant pricing in each consumption segment.
- p_1 and p_2: coefficients representing the penalty percentage for distances from the start and end of the segment, respectively.
- d_s: signifies the distance from the start of the segment.
- d_e: signifies the distance from the end of the segment.

The metric was computed using both mean and weighted mean methods to ensure accuracy and inclusivity in evaluation.

8.2.2 *Application results.*

After examination, both coefficients p_1 and p_2 were assigned as 0.25. Tables 2, 3, and 4 provide an overview of the mean and weighted average of the metric derived from the Low Carbon London dataset, organized by year. These tables offer insights on the algorithms' performance trends over time. Notably, the data highlights the significance of Fuzzy C-Means with Corrections, positioning it as the

Table 2: Results of Metrics for Low Carbon London dataset during 2012

2012	metric		Low	Normal	High
	mean	weighted average	mean		
Heuristic Thresholding	0.628	0.645	0.421	0.675	0.726
KMeans	0.644	0.645	0.641	0.652	0.629
FCM Vanilla	0.624	0.634	0.505	0.679	0.663
FCM Correction 1	0.624	0.634	0.505	0.679	0.663
FCM Correction 2	0.679	0.679	0.508	0.688	0.841
FCM Correction 1 and 2	0.679	0.679	0.508	0.688	0.841

Table 3: Results of Metrics for Low Carbon London dataset during 2013

2013	metric		Low	Normal	High
	mean	weighted average	mean		
Heuristic Thresholding	0.631	0.659	0.331	0.714	0.744
KMeans	0.638	0.634	0.604	0.683	0.627
FCM Vanilla	0.615	0.646	0.394	0.699	0.657
FCM Correction 1	0.615	0.646	0.394	0.699	0.657
FCM Correction 2	0.668	0.685	0.394	0.708	0.834
FCM Correction 1 and 2	0.668	0.685	0.394	0.708	0.839

Table 4: Results of Metrics for Low Carbon London dataset during 2014

2014	metric		Low	Normal	High
	mean	weighted average	mean		
Heuristic Thresholding	0.585	0.613	0.301	0.722	0.651
KMeans	0.608	0.602	0.587	0.686	0.447
FCM Vanilla	0.588	0.618	0.322	0.716	0.639
FCM Correction 1	0.588	0.618	0.322	0.716	0.639
FCM Correction 2	0.636	0.658	0.322	0.732	0.776
FCM Correction 1 and 2	0.637	0.661	0.322	0.732	0.771

best choice for tariff design. Additionally, the analysis underscores the algorithms' proficiency in detecting fluctuations in high tariff periods, showcasing their adaptability to dynamic demand conditions. However, a notable limitation arises in accurately identifying changes in low tariff periods. Nonetheless, Fuzzy C-Means, particularly when combined with Corrections (1, 2, or both), emerge as the optimal performer, consistently achieving the highest accuracy scores.

Tables 5 and 6 illustrate the mean and weighted average of the metric for the CAMSL dataset, delineated by year. Notably, in Table 5, focusing on the weighted average, the analysis for 2017 unveils KMeans as the top performer, followed closely by FCM with Correction 1. Meanwhile, Table 6, addressing also the weighted average, reveals that Heuristic Thresholding demonstrates the most robust performance for 2018, with KMeans exhibiting the subsequent best performance. This trend can be attributed to the distinct segmentation of consumption into discrete segments and the absence of significant "spikes" throughout the day.

9 CONCLUSIONS AND FUTURE WORK

This paper explores three unsupervised learning algorithms applied to dynamic value design, focusing on Fuzzy C-Means and its ability to explain clustering through the membership function. In addition, two heuristic algorithms are introduced to address abnormal periods based on the membership function and to handle multiple pricing changes within a short time.

The study uses five datasets with different consumption patterns in terms of load and the frequency of "valleys" and "peaks" per day to develop and evaluate the algorithms. A new metric was proposed to evaluate the overlap and the delay of tariff adjustments.

Experimental results show that Heuristic Thresholding and KMeans are more effective in identifying tariffs for well-segmented consumption, i.e., without "sharp" peaks. In contrast, Fuzzy C-Means, when combined with corrections 1 and 2, outperforms in identifying tariffs for daily consumption with numerous "valleys" and "peaks", as well as abnormal periods.

Future research should explore the integration of the proposed metric with economic analysis for better refined tariff design. Moreover, evaluating the algorithms using synthetic consumption data with added noise will provide insights into the limits of the grid.

Table 5: Results of Metrics for CAMSL dataset during 2017

2017	metric		Low	Normal	High
	mean	weighted average	mean		
Heuristic Thresholding	0.514	0.492	0.479	0.622	0.529
KMeans	0.519	0.535	0.359	0.673	0.476
FCM Vanilla	0.504	0.531	0.332	0.644	0.463
FCM Correction 1	0.504	0.531	0.331	0.643	0.463
FCM Correction 2	0.503	0.529	0.331	0.645	0.461
FCM Correction 1 and 2	0.503	0.531	0.331	0.644	0.461

Table 6: Results of Metrics for CAMSL dataset during 2018

2018	metric		Low	Normal	High
	mean	weighted average	mean		
Heuristic Thresholding	0.524	0.506	0.501	0.569	0.624
KMeans	0.499	0.488	0.373	0.653	0.491
FCM Vanilla	0.492	0.481	0.357	0.645	0.503
FCM Correction 1	0.492	0.481	0.356	0.646	0.501
FCM Correction 2	0.492	0.481	0.357	0.644	0.502
FCM Correction 1 and 2	0.492	0.481	0.356	0.645	0.499

ACKNOWLEDGMENTS

REFERENCES

[1] Jamshid Aghaei and Mohammad-Iman Alizadeh. 2013. Demand response in smart electricity grids equipped with renewable energy sources: A review. *Renewable and Sustainable Energy Reviews* 18 (2013), 64–72. https://doi.org/10.1016/j.rser.2012.09.019

[2] Howard Anton. 1993. *Elementary Linear Algebra* (7th ed.). Wiley, New Jersey.

[3] BloombergNEF. 2022. Neo 2022. https://about.bnef.com/new-energy-outlook/. Accessed on 12/04/202.

[4] M Charwand, M Gitizadeh, and P Siano. 2017. A new active portfolio risk management for an electricity retailer based on a drawdown risk preference. *Energy* 118 (2017), 387–398.

[5] Valeria Di Cosmo, Sean Lyons, and Anne Nolan. 2014. Estimating the Impact of Time-of-Use Pricing on Irish Electricity Demand. *The Energy Journal* 35, 2 (2014), 117–136. https://doi.org/10.5547/01956574.35.2.6

[6] Samarjit Das. 2013. Pattern Recognition using the Fuzzy c-means Technique. *International Journal of Energy, Information and Communications* 4, 1 (2013), 1–14.

[7] E Dehnavi and H Abdi. 2016. Optimal pricing in time of use demand response by integrating with dynamic economic dispatch problem. *Energy* 109 (2016), 1086–1094.

[8] Arnoud V. den Boer. 2015. Dynamic pricing and learning: Historical origins, current research, and new directions. *Surveys in Operations Research and Management Science* 20, 1 (2015), 1–18. https://doi.org/10.1016/j.sorms.2015.03.001

[9] MS Ding, XY Wang, JW Wang, ZF Yang, HW Zhong, and JL Yang. 2017. A dynamic period partition method for time-of-use pricing with high-penetration renewable energy. In *IEEE Conf. Ene. Int. Ene. Sys. Int. (E12)*. Beijing, China, 1–12. https://doi.org/10.1109/EI2.2017.8245547

[10] Xiaoli Duan, Sanwei Liu, Fuyong Huang, Xiangyu Fan, Jianjia Duan, Zeyu Zeng, Ting Yu, Lipeng Zhong, and Bin Dai. 2023. Peak-Valley Time Period Partition of TOU Tariff Based on Fuzzy Equivalence Relation Clustering Algorithm. *IAENG International Journal of Applied Mathematics* 53, 4 (2023), 1183–1188.

[11] Onur Elma, Akın Taşcıkaraoğlu, A. Tahir İnce, and Uğur S. Selamoğulları. 2017. Implementation of a dynamic energy management system using real time pricing and local renewable energy generation forecasts. *Energy* 134 (2017), 206–220. https://doi.org/10.1016/j.energy.2017.06.011

[12] R García-Bertrand. 2013. Sale prices setting tool for retailers. *IEEE Trans Smart Grid* 4, 4 (2013), 2028–2035.

[13] FF Guo, XX Wang, and J Shen. 2016. Adaptive fuzzy c-means algorithm based on local noise detecting for image segmentation. *IET Image Proc.* 10, 4 (2016), 272–279.

[14] AR Hatami, H Seifi, and MK Sheikh-El-Eslami. 2011. A stochastic-based decision-making framework for an electricity retailer: time-of-use pricing and electricity portfolio optimization. *IEEE Trans Power Syst* 26, 4 (2011), 1808–1816.

[15] Matthias Hofmann and Karen Byskov Lindberg. 2024. Residential demand response and dynamic electricity contracts with hourly prices: A study of Norwegian households during the 2021/22 energy crisis. *Smart Energy* 13 (2024), 100126. https://doi.org/10.1016/j.segy.2023.100126

[16] Zexuan Ji, Yong Xia, Qiang Chen, Quansen Sun, Deshen Xia, and David Dagan Feng. 2012. Fuzzy c-means clustering with weighted image patch for image segmentation. *Applied Soft Computing* 12, 6 (2012), 1659–1667. https://doi.org/10.1016/j.asoc.2012.02.010

[17] Xin Jin and Jiawei Han. 2010. *K-Means Clustering*. Springer US, Boston, MA, 563–564. https://doi.org/10.1007/978-0-387-30164-8_425

[18] L Kaira, M Nthontho, and S Chowdhury. 2014. Achieving demand side management with appliance controller devices. In *2014 49th International Universities Power Engineering Conference (UPEC)*. 1–6. https://doi.org/10.1109/UPEC.2014.6934611

[19] Yohei Kiguchi. 2021. *Residential Demand Response using Electricity Smart Meter Data*. Ph. D. Dissertation. Apollo - University of Cambridge Repository. https://doi.org/10.17863/CAM.79069

[20] Yohei Kiguchi, Melvyn Weeks, and Riku Arakawa. 2020. CAMSL Dataset: Smart Meter Data under Time-Of-Use Tariffs in Japan. Mendeley Data. https://doi.org/10.17632/cmpsyncmmk.1

[21] H Li and L Jia. 2011. Study on peak and valley time periods partition of TOU price. *Power Economy* 9 (2011), 1–3.

[22] R Li, Z Wang, C Gu, F Li, and H Wu. 2016. A novel time-of-use tariff design based on Gaussian mixture model. *Appl Energy* 162 (2016), 1530–1536.

[23] Y. Liu, J. Gou, and Z. Fan. 2019. Educational data mining for students' performance based on fuzzy c-means clustering. *The Journal of Engineering* 2019 (2019), 8245–8250. Issue 11. https://doi.org/10.1049/joe.2019.0938

[24] F Nie, X Zhao, R Wang, X Li, and Z Li. 2020. Fuzzy K-means clustering with discriminative embedding. *IEEE Trans. Knowl. Data Eng.* 34, 3 (2020), 1221–1230.

[25] Open Power System Data. 2020. Open Power System Data: A Free and Open Data Platform for Power System Modelling. https://open-power-system-data.org/.

[26] S Pei, H Chen, F Nie, R Wang, and X Li. 2022. Centerless clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 1 (2022), 167–181.

[27] DF Rogers and GG Polak. 2013. Optimal clustering of time periods for electricity demand-side management. *IEEE Trans Power Syst* 28, 4 (2013), 3842–3851.

[28] J. Schofield, R. Carmichael, S. H. Tindemans, M. Bilton, M. Woolf, and G. Strbac. 2016. Low Carbon London Project: Data from the Dynamic Time-of-Use Electricity Pricing Trial, 2013. UK Data Service. https://doi.org/10.5255/UKDA-SN-7857-1

[29] D Shrinivasan, S Rajgarhia, B Radhakrishnan, A Sharma, and HP Kincha. 2017. Game-Theory based dynamic pricing strategies for demand side management in smart grids. *Energy* 126 (2017), 132–143.

[30] M Soland, S Loosli, J Koch, et al. 2018. Acceptance among residential electricity consumers regarding scenarios of a transformed energy system in Switzerland a focus group study. *Energy Efficiency* 11 (2018), 1673–1688. https://doi.org/10.1007/s12053-017-9548-x

[31] Daiva Stanelyte, Neringa Radziukyniene, and Virginijus Radziukynas. 2022. Overview of Demand-Response Services: A Review. *Energies* 15, 5 (2022). https://doi.org/10.3390/en15051659

[32] R. Suganya and R Mercy Shanthi. 2012. Fuzzy C- Means Algorithm- A Review. https://api.semanticscholar.org/CorpusID:17666771

[33] Peng Wang, Yiwei Ma, Zhiqi Ling, and Genhong Luo. 2023. Peak-valley period partition and abnormal time correction for time-of-use tariffs under daily load curves based on improved fuzzy c-means. *IET Generation, Transmission & Distribution* 17, 24 (2023), 5396–5409. https://doi.org/10.1049/gtd2.13052 arXiv:https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/gtd2.13052

[34] YJ Wang, YK Sun, and SQ Han. 2020. Research on the time division method of peak-valley time-of-use tariff. *Management & Technology of SME* 9 (2020), 84–87.

[35] P Yang, G Tang, and A Nehorai. 2013. A game-theoretic approach for optimal time-of-use electricity pricing. *IEEE Trans Power Syst* 28, 2 (2013), 884–892.

[36] C Zheng, Y Tian, and YH Chen. 2016. Research of peak and valley time period partition based on fuzzy clustering. *Electrical Engineering* 6 (2016), 13–17.

Received 18 May 2024