Expert Systems With Applications 227 (2023) 120123

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Electricity consumption forecasting with outliers handling based on clustering and deep learning with application to the Algerian market

Dalil Hadjout[a], Abderrazak Sebaa[b,c], José F. Torres[a], Francisco Martínez-Álvarez[a,*]

[a] Data Science and Big Data Lab, Pablo de Olavide University, ES 41013 Seville, Spain

[b] Laboratoire LITAN, Ecole Supérieure en Sciences et Technologies de l'Informatique et du Numérique, RN 75, Amizour, 06300 Bejaia, Algeria

[c] LIMED Laboratory, Faculty of Exact Sciences, University of Bejaia, 06000 Bejaia, Algeria

ARTICLE INFO

Keywords:
Time series forecasting
Deep learning
Clustering
Outlier detection
Electricity consumption

ABSTRACT

The reduction of electricity loss and the effective management of electricity demand are vital operations for production and distribution electricity enterprises. To achieve these goals, accurate forecasts of aggregate and individual electricity consumers are necessary. A novel multistep forecasting method is developed to forecast medium-term electricity consumption of the Algerian economic sector. The proposed method goes through the following three steps: cleaning steps, clustering steps and forecasting step of each cluster. The aim of the first step is to detect and then replace outliers. To complete the first phase, Robust Exponential and Holt-Winters Smoothing algorithms are adapted. Then, to carry out accurate forecasting at a lowest level, K-Shape and K-Means clustering methods are utilized to extract similarities and identify customer consumption patterns as a second step. The third step entails developing a deep learning model based on Gated Recurrent Units to forecast the electricity consumption in each cluster. To validate the proposed method, we compared our results to the most known methods in literature like Autoregressive Integrated Moving Average, Seasonal Grey Model, LSTM networks, Temporal Convolutional Networks and two ensemble models. The results of several experiments conducted with 2000 electricity consumers during 14 years from an Algeria province (Bejaia) demonstrate that the proposed method provides remarkable prediction performances. Thus, prediction performances of the K-Shape-based clustering method reach much higher prediction accuracy. According to the MAPE metric, the results of the best predictions are equal to 2.04%. It is also notable that 87% of the clients have a considerably low prediction error.

1. Introduction

To adequately plan electricity production and to avoid disturbance and loss of electricity supply, it is necessary to have accurate forecasts of aggregate and individual electricity consumers. The world's economies are significantly impacted by the prediction of electricity consumption. Thus, Non-technical loss (NTL), also known as commercial loss, is a significant issue for electricity distribution societies (Gutiérrez-Avilés et al., 2018). Energy that is provided but not invoiced to the consumer due to billing or metering mistakes or dishonest customers is known as NTL. Massive losses are caused by NTL. A customer's monthly consumption prediction is a strategy to avoid erratic or unbillable use brought on by field meter system failures.

The main goal of our proposed method is to enrich the literature by investigating for the first time Algerian electricity and address the challenging problem of multiple times series analysis applications in the context of the supplier electricity companies. Indeed, this study represents the first of its kind in Algeria. The country's market data shows that it is a promising and strategic market, situated at the center of Europe, Africa, and the Middle East. Algeria is a significant exporter of energy, and thus forecasting electricity consumption is not only crucial for the country itself but also for other countries in the region. In this context, the prediction study of Algerian electrical data becomes more than necessary, as it can offer valuable insights into energy consumption patterns, which can inform energy policies and strategies across the region.

On the one hand, our proposed solution offers practical applicability in the real field of the company by enabling sales analysis, reducing non-technical losses (NTL), and driving the operational team for ground intervention. The solution can optimize interventions in the field, leading to significant benefits such as reduced costs and increased efficiency. Additionally, having only a few models to maintain simplifies the overall maintenance process, making it easier for the company

* Corresponding author.
E-mail addresses: dhad@alu.upo.es (D. Hadjout), abderrazak.sebaa@univ-bejaia.dz (A. Sebaa), jftormal@upo.es (J.F. Torres), fmaralv@upo.es (F. Martínez-Álvarez).

https://doi.org/10.1016/j.eswa.2023.120123
Received 21 September 2022; Received in revised form 6 April 2023; Accepted 9 April 2023
Available online 20 April 2023

to manage and update its forecasting system. Overall, our solution has the potential to offer significant value to the company by improving operational efficiency and reducing NTL.

On other hand, the review of deep learning shows that one of the drawbacks of currently available deep learning models is their lack of consideration for multiple time series. However, in reality, electricity consumption data corresponds to several customers, with each customer representing a separate time series source that differs from the others. It is crucial for distribution companies to understand the profile of each customer as it enables efficient sales analysis, which can directly impact loss reduction. Therefore, the need to master each customer's consumption profile remains very valuable, and addressing the challenge of multiple time series is essential for developing accurate and effective deep-learning models for electricity consumption forecasting. In addition, very little research has been identified that deals with the prediction of consumption in the economic sector. However, economic customers represent an important portion of global consumption. Knowing the consumption profile of this type of customer allows us to optimize sales while also anticipating the future global energy demand.

Indeed, predicting High Voltage type A (HVA) consumers' medium-term electricity consumption is a problem that we aim to address. To achieve this goal, cluster prediction (Scitovski, Sabo, Martínez-Álvarez, & Ungar, 2021) and deep learning are being investigated in order to forecast electricity more accurately on a global and individual granularity (Pérez-Chacón, Luna-Romera, Troncoso, Martínez-Álvarez, & Riquelme, 2018). Clustering can uncover patterns and structures in the data that may not be immediately apparent when looking at individual data points. By grouping similar data points together, clustering can identify trends, relationships, and outliers in the data. In our experimental data, HVA customers comprise different categories of users from various economic sectors, especially the industry, and services.

The proposed approach consists of three main steps. First, we apply a preprocessing data method called Robust Exponential and Holt-Winters Smoothing (REHWS) in order to detect and replace outliers. In the second step, we cluster the dataset's clients according to their level of similarity. We apply K-Means and K-Shape algorithms, two different types of clustering techniques, to obtain customer clusters. The proposed methods rely on clustering, which is a technique used to group similar data points together based on their features or characteristics.

Then, in the third step, we use direct forecasting to extract intricate patterns from sequence data using a deep learning (DL) Gated Recurrent Unit (GRU) model and a Bayesian Optimizer (BO). Clearly, for each cluster model, the best hyper-parameters are obtained using the Bayesian optimizer.

Real customer data from Algeria has been utilized to validate our methodology. Sonelgaz, Algeria's top electricity and gas distributor, has given these statistics. It provides gas to 6 million users and electricity to 10 million. Additionally, Sonelgaz has a number of companies that are in charge of producing, transmitting, and distributing energy.

The comparisons carried out of the proposed method with other well-established techniques have shown a significant rise in accuracy. Therefore, to validate the clustering approach, we compare our results to different types of existing methods in the field of electricity consumption prediction research. Seasonal Autoregressive Integrated Moving Average (SARIMA) has been chosen as a statistical approach, Seasonal Grey Model (SGM) as a mathematical method, Temporal Convolutional Networks (TCN) and GRU, Long Short-Term Model (LSTM) as deep learning models, and with two recent ensemble methods.

The contributions of this work are significant on various levels, as it addresses an important issue in Algeria regarding electricity consumption forecasting, and it extends to economic, technical, and modeling levels, making it a valuable contribution to the field of energy management and forecasting.

At the economic level, this is the first time that the issue of electricity consumption forecasting has been addressed in Algeria using a three-step process that enhances the standard method with anomaly detection, clustering, and hyper-parameter optimization. Our proposed solution is highly accurate and applicable in the real world of distribution companies. It addresses a multiple time series dataset containing real information about Algerian customers in the economic category. Although this category represents less than 1% of the total number of consumers, they account for more than 45% of the total electricity consumption in Algeria. Their electricity consumption patterns are heterogeneous, and the variation of consumption in time between consumers is vast. Furthermore, the demand for electricity in this category is growing rapidly, making it challenging and interesting to study this category.

At the modeling level, since we are dealing with multiple time series, we have adapted each standard technique proposed to suit this context. Specifically, we have developed an algorithm to adjust the robust anomaly detection technique in the context of multiple time series. Predicting multiple time series is a challenging task due to several reasons. Firstly, the non-linearity of the series does not allow for simple mathematical modeling. Additionally, outliers may occur due to index reading and data entry errors, making it difficult to accurately predict consumption patterns. Furthermore, the length of the series varies from one customer to another, as does the consumption scale. These differences complicate the modeling process, as a single model may not accurately capture the diverse consumption patterns of each customer. Finally, empty values may also exist in the data, adding to the difficulty of accurately modeling the time series. Despite these complexities, our proposed framework provides accurate predictions of electricity consumption patterns that can be invaluable to distribution companies seeking to optimize their energy management strategies.

At the technical level, we have modified the Robust Exponential and Holt-Winters Smoothing methods to use them in the context of multiple time series. We have adapted two different clustering techniques, K-Means and K-Shape, to handle multiple sources of time series. We have also developed a powerful Bayesian optimizer to ensure that our model is optimally trained for hyper-parameter tuning.

To provide a clear overview of the contributions made by this work, the following paragraph presents a summary of the paper's significant achievements. These contributions represent valuable advancements in the field of energy management and forecasting, and address challenges related to electricity consumption forecasting in Algeria. Below is a list of the key contributions made by this work:

1. We investigated two clustering methods by grouping customers with structural similarity in their consumption over time.
2. We have added a cleaning phase to manage outliers using REHWS.
3. We developed a deep GRU model as a forecasting model.
4. We implemented a BO to get the model's top hyper-parameters.
5. The superiority of the proposed model was shown by a fair performance comparison with several well-established methods.
6. Global and individual forecasting are included in sale analysis software.

The rest of this paper is structured as follows: Section 2 examines a similar paper on electricity consumption-based clustering prediction and highlights the need for research by work that targets individual prediction. Section 3 explains the suggested methodology. Section 4 presents the experimental aspects of our work. Section 5 highlights the most meaningful results (global and individual) of Bejaia's electricity consumption forecasts (Algeria). Section 6 presents the outcomes of this research.

2. Related work

We review some recent studies on the prediction of electricity consumption. There are many different prediction techniques, but recently, deep learning-based methods have become more and more the

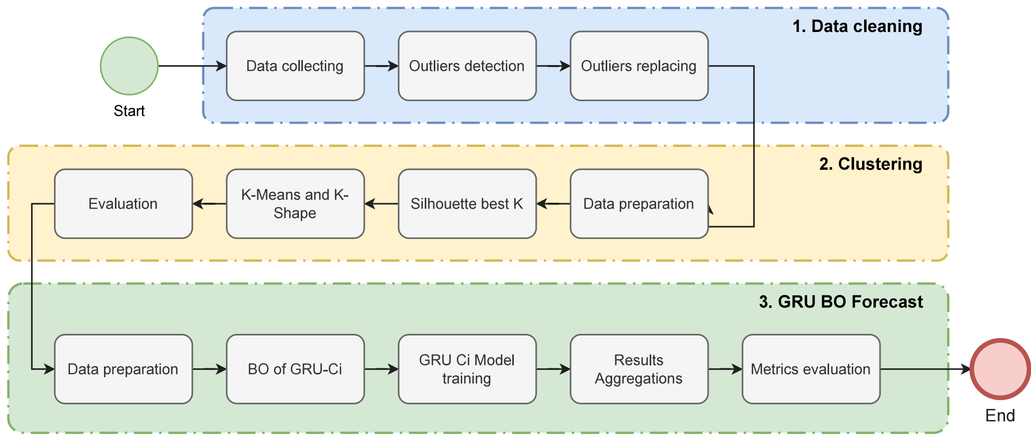

Fig. 1. Methodology description.

focus of research (Torres, Hadjout, Sebaa, Martínez-Álvarez, & Troncoso, 2021). For illustration, standard deep learning models like deep feed forward neural network (DFFNN) (Torres, Galicia, Troncoso, & Martínez-Álvarez, 2018), LSTM (Nakyoung, Minkyung, & Kyun, 2018), GRU (Bingchun, Chuanchuan, Arlene, & Quan, 2017), TCN (Lara-Benítez, Carranza-García, Luna-Romera, & Riquelme, 2020), methods for ensemble deep learning that incorporate many DL models (Divina, Torres, García-Torres, Martínez-Álvarez, & Troncoso, 2020) have been proposed in the literature during the last years.

Some papers have also been published discussing how to predict electricity consumption using clustering methods. Thus, Torabi, Hashemi, Saybani, Shamshirband, and Mosavi (2019) introduced a method based on the idea of clustering and the Artificial neural network (ANN) and Support-Vector Machine (SVM) model as an approach to increase the accuracy of the load forecast. By combining the K-Shape technique with deep learning techniques like Restricted Boltzmann Machines and Deep Belief Networks, the works presented in Fahiman, Erfani, Rajasegarar, Palaniswami, and Leckie (2017) and Yang, Ning, Deb, Zhang, Cheong, et al. (2017) proposed a method to improve the prediction of electricity consumption. In Jahangir et al. (2020) a prediction method based on Bi-directional LSTM networks and micro-clustering is proposed. The research Laurinec (2018) developed a prediction approach based on the influence of time series representation and clustering.

The global prediction is well represented in the literature, and the results obtained are also very powerful and very accurate, especially with the advent of deep learning methods. By contrast, only a few efforts have been made towards individual forecasting in recent years, mainly focusing on deep learning methods LSTM (Kong et al., 2017; Torres, Martínez-Álvarez, & Troncoso, 2022) and pooling deep Recurrent Neural Network (RNN). In Gholizadeh and Musilek (2022), the effectiveness of federated learning is evaluated for both the aggregate load and the short-term forecast of individual housing loads. The authors discussed the advantages and disadvantages of this method by comparing it with centralized and local learning schemes. The authors in paper (Sepulveda, Diniz, Diniz, et al., 2021) proposed an optimized gradient boosting regressor for a rapid parameters optimization. A modified version of the Particle Swarm optimization is used. There are few papers in the literature that use ensemble learning as a strategy for predicting electricity consumption. Recently, in Hadjout, Torres, Troncoso, Sebaa, and Martínez-Álvarez (2022), by training three separate deep learning models, LSTM, GRU, and TCN, with the electric consumption of the economic sector clients, the authors suggested a deep learning ensemble. The authors used two distinct types of approach, static and dynamic ensembles, to extrapolate predictions from trained models. These methods allowed for the implementation of combinations depending on the weights of the various models, which resulted in a more effective ensemble configuration. Last, very few articles have dealt with the prediction of individual electricity consumption of economic customers (Bai et al., 2021).

In fact, our paper aims to investigate the issue of individual prediction of economic customers. The proposed approach targets both the global prediction by the aggregation strategy and the individual prediction by learning in depth the pattern contained in the data for each customer.

3. Methodology

The organization of this section is as follows. In Section 3.1, we first establish the problem formulation. In Section 3.2, we provide a summary of the theoretical requirements for the suggested strategy. The approach suggested to forecast electricity use is described in Section 3.3.

3.1. Mathematical formulation of the problem

Our problem is a one-step univariate time series prediction (prediction horizon equal to 1, $h = 1$) using a window of W previous values. The Eq. (1) formulates this:

$$\hat{y}_{t+1} = f\left(y_t, y_{t-1}, \dots, y_{t-(W-1)}\right), \tag{1}$$

where the aim is to retrieve the complex function f, y_t is the target time series and $\hat{y}_{t+1}$ the estimated value for actual y_{t+1}.

A forecast of N customers' consumption for the upcoming month is required in this case. Let $Cit + 1$ be the ith customer's consumption at time $t = 1$ and $hatCit + 1$ be the expected cost of that consumption. Therefore, as indicated in Eq. (2), the objective is to correctly estimate the value at $t = 1$ for each of the N customers.

$$\hat{C}_{t+1} = \sum_{i=1}^{N} \hat{C}^{i}_{t+1} \tag{2}$$

where $\hat{C}(t+1)$ represents the forecasting of the consumption for the following value of all N clients.

In our approach, a forecast is made for each group and the predictions for each cluster are added up after consumers are first separated into various clusters, the load consumption for each cluster is added, and finally a forecast is made for each group.

3.2. Theoretical background

In this section, we recall the essential theoretical concepts used in our work, namely, time series forecasting, time series clustering, similarity measures, Gated Recurrent Units, Bayesian optimization and outliers detection and replacement.

3.2.1. Time series forecasting

To predict a time series' future values is one of the analysis's most crucial goals. There are two principal classes in the literature: Box-Jenkins (Box & Jenkins, 2008) and machine learning-based models (Martínez-Álvarez, Troncoso, Asencio-Cortés, & Riquelme, 2015).

During the forecasting process, the first question to address is how to describe time series data. Overall, we note a variety of forms and different stochastic processes that can be represented by time series forecasting models. First, the moving average (MA) and auto regressive (ARMA) models of linear time series are often employed in literature. Moreover, research has led to combining these two lasts giving the Autoregressive Moving Average model (ARMA) and the Autoregressive Integrated Moving Average model (ARIMA). Then, the Auto-Regressive Fractionally Integrated Moving Average model (ARFIMA) generalizes ARMA and ARIMA models. For seasonal time series forecasting, a variation of ARIMA is used, the SARIMA model is used (Chikobvu & Sigauke, 2012).

Research has also suggested the SGM as a forecasting method. It can be used to forecast energy consumption using grey theory, and it can be built for at least four data points or ambiguity data (Lee & Tong, 2011).

Various architectures of deep learning have been published in the literature, including Adversarial Networks (AN), Convolutional Neural Networks (CNN), GRU, DFFNN, and RNN.

The problem's characteristics to be addressed will determine which architecture is used. A thorough analysis of the various models has been provided in Torres et al. (2021) and a practical analysis of these systems is available in Lara-Benítez, Carranza-García, and Riquelme (2021). There are several prediction strategies reported in the literature (Fahiman et al., 2017):

1. Centralized forecasting.
2. Distributed forecasting.
3. Clustering-based forecasting.

Ensemble learning is another method for forecasting electricity. By reducing their bias and variance, multiple learning algorithms are combined to increase the prediction accuracy of each individual model (Hadjout et al., 2022).

3.2.2. Time series clustering

A set of unlabeled time series is classified into groups using time series clustering, where each sequence that forms a cluster should be similar or coherent. It might operate at the shape-level when applied to numerous single time series or at the structure-level when applied to a individual lengthy time series. The large dimensional, chronological order, and data noise are the main problems with time series clustering. Time series clustering can be temporal-proximity-based (direct use of frequency or time domain raw data), representation-based (using features taken from the raw data indirectly), or model-based (using a model created from the raw data).

The K-Shape time-series clustering algorithm is designed to create well-defined groups of time-series data with consistent shapes. Unlike traditional clustering methods such as K-means, which treat individual observations within a time series as independent attributes, K-shape considers the shape of the time series when clustering. This approach uses a domain-independent form of cross-correlation as the distance measure between time series (Paparrizos & Gravano, 2015). By doing so, K-shape is able to efficiently and effectively derive a shape-based distance measure for comparing time series. This enables the algorithm to create clusters of time series with similar shapes, providing a more meaningful clustering solution for time-series data (Fahiman et al., 2017).

3.2.3. Similarity measures

In the literature, there are more than a dozen different distance measures for comparing time series data. Euclidean distance (ED), Dynamic Time Warping (DTW), and Bilateral Shape-Based Distance (SBD) stand out among them. It is possible to think of DTW as an ED extension that provides a local (non-linear) alignment. The most popular algorithm for comparing time series with the best performance among other metrics is DTW (Yang et al., 2017). As an alternative, Shape-Based Distance (SBD) is based on the statistical measure of cross-correlation that allows us to assess how similar two sequences are to one another (Paparrizos & Gravano, 2015).

3.2.4. Gated recurrent units

GRU is RNN type model and first appeared in 2014 (Chung, Gulcehre, Cho, & Bengio, 2014) as an LSTM simplification. One of the most often used variants that scientists have chosen is GRU, which they have discovered to be trustworthy and useful for a variety of problems. The use of gates in RNNs has enhanced the capture of extremely long-range dependencies. GRU is a simpler model and computationally quicker since it only has two gates, however, LSTM is more strong and effective because it contains three gates instead of two, the Γ^r relevance gate and Γ^u update gate. The Γ^u gate will decide whether the c_t memory state is or is not updated using the $\widetilde{c}_t$ memory state candidate. The Γ^r gate establishes the relevance of c_{t-1} in computing $\widetilde{c}_t$, the following candidate for c_t. According to the following equations, a GRU unit is:

$$\Gamma^u = \sigma(W_u[c_{t-1}, x_t] + b_u) \tag{3}$$

$$\Gamma^r = \sigma(W_r[c_{t-1}, x_t] + b_r) \tag{4}$$

$$\widetilde{c}_t = \tanh(W_c[\Gamma^r * c_{t-1}, x_t] + b_c) \tag{5}$$

$$c_t = \Gamma^u * \widetilde{c}_t + (1 - \Gamma^u) * c_{t-1} \tag{6}$$

$$a_t = c_t \tag{7}$$

With W_r and W_u, and b_r and b_u are the weights and bias that are responsible for the behavior of the gates Γ_r and Γ_u, respectively, and b_c and W_c are the bias and weights of the candidate memory cell $\widetilde{c}_t$.

3.2.5. Hyper-parameter Bayesian optimization

An automated solution is required to prevent spending hours on the hyper-parameter optimization phase. BO is becoming a potent method for deep learning architecture optimization (Cho et al., 2020).

The Bayesian theorem is the foundation of Bayesian optimization. In order to update the posterior of the optimization function, it sets a prior over the optimization function and collects data from the previous sample (Wu et al., 2019).

3.2.6. Detecting and replacing outliers

Data preprocessing plays an important role in electricity forecasting. In time series analysis, predicting future outliers is crucial for minimizing the average prediction error (Jeenanunta, Abeyrathna, Dilhani, Hnin, & Phyo, 2018).

Our exploration of the literature has brought to light several ways of classifying methods for detecting obstacles. Three complementary approaches (Blázquez-García, Conde, Mori, & Lozano, 2022; Braei & Wagner, 2020; Gupta, Gao, Aggarwal, & Han, 2013) describe ways of classifying different methods. Indeed, the authors proposed three types of methods of outlier detection model base, density-based and histogramming based methods in Blázquez-García et al. (2022). Another classification method was proposed in Gupta et al. (2013), which included prediction models profile, Deviants and similarity-based techniques (which keep a typical profile and compare a new time point to this profile to determine if it is an outlier). By setting the deviation rather than the space as in traditional models, these information theoretic models investigate the trade-off between space and deviation. We concur that there are three primary groups into which all anomaly detection techniques on time series data may be categorized (Braei & Wagner, 2020):

1. Statistical methods.
2. Traditional machine learning methods.
3. Deep learning methods.

3.2.7. Performance assessment

The suggested deep learning models' performance is assessed using Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). Moreover, in order to evaluate the performance of our approach at all levels, we introduce the symmetric mean absolute percentage error (sMAPE), which normalizes the relative errors by dividing them by both the actual and predicted values. The sMAPE is a variation of the MAPE that uses half of the sum of the actual and forecast values as the divisor (Kim & Kim, 2016). The MAPE metric can lead to large errors, particularly in cases where the actual values are close to zero, leading to asymmetric errors. The sMAPE metric addresses this issue by computing the errors as a percentage of the sum of the actual and predicted values, resulting in a symmetric error measure that ranges between 0% and 100%. This can be particularly helpful when evaluating individual predictions as it allows for a more accurate and fair comparison across different customers, regardless of their actual consumption levels.

Indeed, for cases where the actual and predicted values are already symmetric, the sMAPE and MAPE metrics will provide the same evaluation of the prediction. This is because sMAPE is almost equal to half of MAPE when the values are symmetric. However, in cases where the actual and predicted values are not symmetric, such as when the actual values are close to zero, sMAPE may provide a more accurate evaluation of the prediction compared to MAPE. Therefore, it is important to consider both metrics when evaluating the performance of the prediction model.

They are mentioned in Eqs. (8), (9), and (10) respectively:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|p_i - a_i| \tag{8}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|p_i - a_i|}{a_i}\cdot 100 \tag{9}$$

$$sMAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|p_i - a_i|}{|p_i| + |a_i|}\cdot 100 \tag{10}$$

where N, p_i and a_i denote respectively the number of samples, the predicted values and the actual values at time i.

3.3. Methodology description

This section explains the strategy suggested to predict time series of electricity consumption. It is made up of three main phases which are described below.

Fig. 1 introduces the methodology flow diagram, where GRU-Ci identifies the application of GRU to the ith cluster generated by K-Means or K-Shape. Each of these steps is described in subsequent subsections.

3.3.1. Clustering phase

To use the dataset in the clustering stage, the dataset must first be preprocessed. Therefore, Eq. (11) illustrates the dataset as a list:

$$D = \{X_1^1, X_2^1, \dots, X_{n1}^1\}, \{X_1^2, X_2^2, \dots, X_{n2}^2\}, \dots, \{X_1^m, X_2^m, \dots, X_{nm}^m\} \tag{11}$$

where x is the monthly electricity consumption value of a customer, n is the number of customers, and m is the size of the training set.

In addition, all input variables are normalized by scaling them to a range of $[-3, +3]$ using the z-score method, shown in Eq. (12):

$$Z = \frac{Z - \mu}{\sigma} \tag{12}$$

where $sigma$ is the standard deviation of the normal variable X, mu is the mean of the normal random variable X and x is a normal random variable.

Clustering algorithms are used after preprocessing. Consumers with comparable consumption time series can be clustered together using clustering, which makes the global noise in each cluster tend to reduce to almost zero. The most popular clustering techniques for time series clustering are K-Means and K-Shape.

The Silhouette index is used to determine the ideal number of clusters, and the DTW and BSD distance measures, respectively, are used for K-Means and K-Shape, respectively, in the clustering process. As an assessment metric, SSE, or sum of square errors, is employed. SSE is the product of the squared deviations between each value and the mean of its cluster. In Eq. (13), the SSE formula is explained :

$$SSE = \sum_{i=o}^{n}(X_i - \bar{X})^2 \tag{13}$$

where n is the total number of samples, X_i is the ith value and $barX$ is the cluster mean.

3.3.2. Anomalies detection and replacing phase

To minimize the effect of outliers on the quality of the prediction as evidenced in the literature, we proceeded to an additional step in the data preparation process. This step consists of detecting the outliers and replacing them. To do this we have adapted REHWS the method proposed in the article (Bartos, Mullapudi, & Troutman, 2019). This technique is chosen for its ease of implementation and ease of determining the parameters and the time series results that were achieved with trend and seasonality. In addition this method proposes a replacement value in the outliers. Furthermore, a wide family of robust statistical techniques exists (Maronna, 2021) that deal with outliers and, in particular, suggest strategies to find them in the datasets that are being analyzed (Martínez-Álvarez, Troncoso, Riquelme, & Aguilar-Ruiz, 2011).

Without the need to develop a parametric model, the simple recursive approach of exponential smoothing may be utilized to predict and smooth a time series. For seasonal and trending time series, the Holt-Winters approach, commonly known as an extension of exponential smoothing known as double exponential smoothing. The authors suggested robust iterations of the Holt-Winters and exponential smoothing methods. The latter are built around a pre-cleaning technique that locates and reduces the weight of outlying data. The robust recursive approach is fairly simple to construct and uses the common smoothing techniques for the precleaned data.

The formal description of the cleaning phase is calculated according the exponential smoothing method. Thus, for a time series y_t, which is observed at $t = 1, \dots, T$. The smoothed series is calculated using the exponential smoothing approach in accordance with Eq. (14):

$$\tilde{y}_t = \lambda y_t^* + (1 - \lambda)\tilde{y}_{t-1} \tag{14}$$

where y_t^* is the cleaned version of observed y_t, $\lambda \in [0, 1]$ is the smoothing parameter, which regulates the degree of smoothing. For both the exponential and Holt-Winters smoothing, the cleaned series y_t^* is calculated according to Eq. (15):

$$y_t^* = \psi\left(\frac{y_t - \hat{y}_{t|t-1}}{\hat{\sigma}_t}\right)\hat{\sigma}_t + \hat{y}_{t|t-1} \tag{15}$$

An H-step-ahead prediction at time t is denoted by $\hat{y}_{t|t-1}$ and the scale of the forecast errors is estimated by $\hat{\sigma}_t$. The ψ function reduces the influence of outlying observations. In this work, we take the Huber function, which is defined in Eq. (16):

$$\psi(x) = \begin{cases} x & if\ |x| < k \\ sign(x)k & otherwise \end{cases} \tag{16}$$

A boundary value based on k should be used in place of the observation if the median absolute deviation (MAD) between the anticipated and observed values is too high. Then, σ is calculated using Eq. (17).

$$\hat{\sigma}_t^2 = \lambda_\sigma p\left(\frac{r_t}{\hat{\sigma}_{t-1}}\right)\hat{\sigma}_{t-1}^2 + (1 - \lambda_\sigma)\hat{\sigma}_{t-1}^2 \tag{17}$$

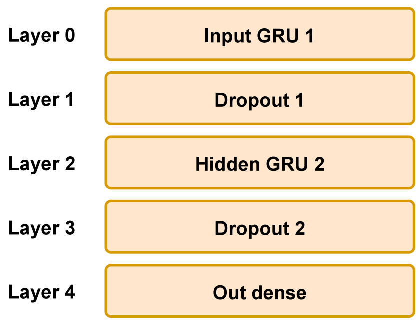

Fig. 2. Deep GRU model architecture.

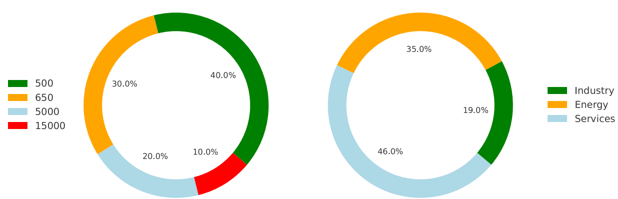

Fig. 3. MPD and activity sector distribution.

where r_t implicitly assumes a normal distribution of forecast errors in one step and is calculated using Eq. (18):

$$r_t = y_t - \hat{y}_{t|t-1} \tag{18}$$

Finally, the p function is the biweight loss function and is calculated using Eq. (19).

$$p(x) = \begin{cases} c_k(1-(1-(x/k)^2)^3) & if\ |x| < k \\ c_k & otherwise \end{cases} \tag{19}$$

where c_k is a constant to achieve the consistency of the scale parameter for a normal error distribution.

In our context, for each customer, the consumption values processed by the robust model as well as any outliers that exceed the predefined MAPE are replaced. The cleaned data will be prepared for the prediction via the clustering-based system.

For each time series of a client, we have proceeded to the detection of anomalies and then to the replacements by values proposed by the REHWS method.

3.3.3. Deep forecasting phase

Next, a BO algorithm based on the GRU model was created with the goal of obtaining the best model for each cluster. The best hyper-parameters for each of the models are provided by this model.

The phase of prediction comes next. This stage entails forecasting electricity consumption using the GRU-Ci model that has been optimized. We proceed to training the model with data from the relevant cluster using the improved GRU models. The consumption of the testing data is then predicted using the trained model.

In Fig. 2 we highlight the layered architecture of the GRU model that has been implemented in our work. The stacked model is described below:

1. Layer 0: GRU layer. It is used as the input layer of the model, with the number of units set to 12, corresponding to the window of inputs (W = 12).
2. Layer 1: First Dropout layer. It is used as a method to reduce the vanishing gradient problem in recurrent models, with a dropout rate ranging between 0.0 and 0.2.
3. Layer 2: Hidden GRU layer. It is used as the first hidden layer, with the number of units ranging from 8 to 64. The activation functions used can be Dense, ReLU, or Sigmoid.

Table 1
Hyper-parameters of the proposed model.

Parameter	Setting
Optimizer	Adam
Loss function	MSE
Metrics	Accuracy, MSE
Epochs	50
Batch	1024
Window	12
Prediction horizon	1

Table 2
BO search space.

Parameter	Search space
Number of GRU units	min = 3, max = 64
Activation function	Dense, ReLU, Sigmoid
Dropout	min = 0.0, max = 0.2, step = 0.01

Table 3
BO settings.

Parameter	Value
Objective	MSE
Maximum trial	10
Seed	42
Executions per trial	2

4. Layer 3: Second Dropout layer. It serves the same purpose as the first dropout layer, with the same dropout rates.
5. Layer 4: Last Dense layer. It is used to compute the final predicted value.

In Table 1 the parameters of the training model in question are shown. In Table 2 the configuration of the search space as well as the parameters of the BO engine are shown. Finally, Table 3 shows the BO parameter setting.

The aggregation step is then completed. In this stage, we will move on from each cluster's forecast to the aggregated prediction data to obtain the overall prediction and the customer level prediction.

4. Experiments

First, Section 4.1 describes the dataset that was utilized. Then, the DCS is presented in Section 4.2. Afterwards, Section 4.3 explains the setup of experiments.

4.1. Dataset description

A time series of energy consumption data was gathered for this study from the Bejaia concession of the Algerian Electricity and Gas Company (SONELGAZ). It covers 13 years (from 2006 to 2019), has 285 432 measures, 1699 users from various economic sectors, and monthly frequency measures.

The consumption data correspond to very different customers in terms of consumption profiles and power requirements. The maximum power demand (MPD) defines the customer's template. This makes the prediction more complex and requires a solution that takes into account the disparity between customers and the difference in consumption scale. Fig. 3 depicts the MPD and activity sector distribution.

4.2. Data collecting system

Sonelgaz uses a DCS (Data collecting system) to measure the electricity consumption of HTA customers. Each customer's meter is equipped with a chip that allows the DCS to collect various electrical

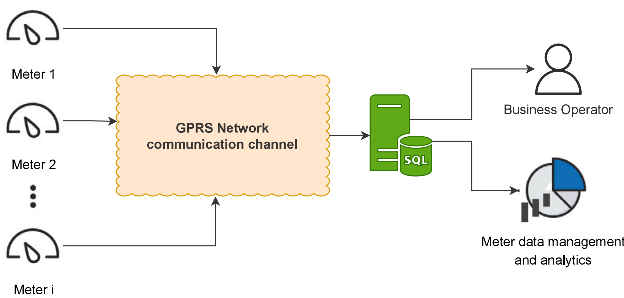

Fig. 4. Sonelgaz's data collecting system.

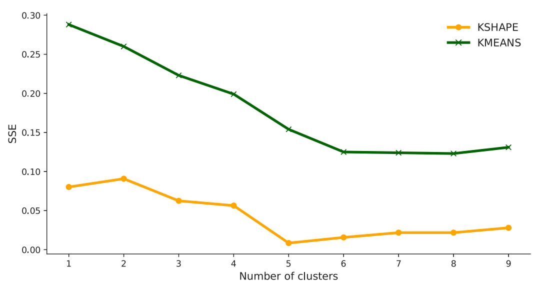

Fig. 5. Best K using Silhouette for K-shape and K-means.

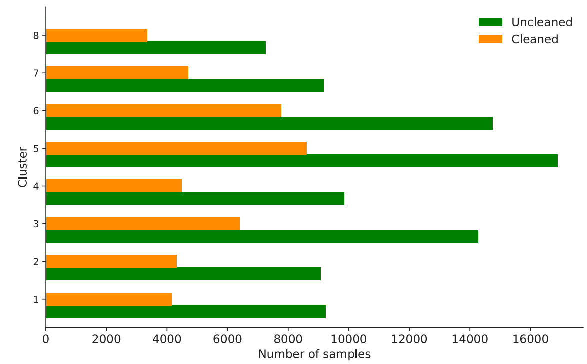

Fig. 6. Cleaned samples, distributed per cluster.

Table 4
Results of clustering: number of clients per cluster for K-Means and K-Shape.

Cluster	K-Shape	K-Means
1	173	61
2	301	200
3	368	127
4	200	274
5	132	114
6	–	66
7	–	264
8	–	68

Table 5
Number of samples cleaned, distributed per cluster.

Cluster	Uncleaned	Cleaned	Rate
1	9234	4146	30.99%
2	9062	4318	32.27%
3	14 268	6399	30.96%
4	9846	4482	31.28%
5	16 896	8605	33.74%
6	14 742	7770	34.51%
7	9173	4699	33.87%
8	7257	3339	31.51%
Global	90 478	43 758	32.60%

consumption data. Data are indexes, loads, or alarms and are collected remotely via a GPRS network. Fig. 4 illustrates the DCS architecture.

In general, DCS improves the quality of index readings, but anomalies may arise that hinder the quality of the data collected and we can mention (Hadjout et al., 2022): meter failures, lack of GPRS network coverage and customer fraud.

4.3. Experimental setting

It is programmed to forecast electricity use for the following month using data from the last 12 months. This indicates that the dataset will have 13 columns after pre-processing, with the first 12 serving as predictions for the 13th column.

K-Means and K-Shape are two clustering techniques that have been utilized to produce the clusters. Once applied the clustering algorithms, K clusters $C = C_1, C_2, \dots, C_K$ are obtained, where each cluster C_i of clients can be used to train a specific deep model.

The Keras 2.3.1 API, the TensorFlow 2.1.0 framework, and Python 3.7 were used to implement each deep learning algorithm. We used Keras Tuner's implementation of Bayesian optimization. Additionally, we employed the Python 3.5+ cross-platform software package Tslearn v0.3.1. In addition to specific clustering models, classification, and regression, Tslearn is machine learning library for time series API for feature extraction and pre-processing (Tavenard, Faouzi, Vandewiele, Divo, et al., 2020).

All experiments were conducted on an Intel Core i5-4310U processor clocked at 2.64 GHz with 3 MB of cache, 2 cores with 4 threads, and 8 GB of RAM, running on a 64-bit Windows operating system.

5. Results

This section summarizes the outcomes of the proposed approach. Results of the clustering step are illustrated in Section 5.1. The data cleaning results are reported in Section 5.2. Section 5.3 describes the prediction results.

5.1. Clustering process

We calculated the appropriate number of clusters based on the clustering findings using the Silhouette approach for both K-Shape and K-means. The optimal K is 5 with an SSE of 0.01 for K-Shape and K is 8 with an SSE of 0.12 for K-Means, as per the findings, as shown in Fig. 5. The number of clients included in each cluster is listed in Table 4.

5.2. Results of cleaning data

On the one hand, Table 5 highlights the distribution of anomalies by clusters, that is, the number of instances that needed to be cleaned. It can be appreciated that there is a balance between all the clusters in terms of anomaly rate. On the other hand, Table 6 indicates the amount of energy resulting after the cleaning process, expressed in GWh.

The result of the cleaning phase shows that 32.60% of the consumption values are corrected by the REHWS method, as shown in Table 5 and in Fig. 6, which corresponds in terms of energy only to 0.75% of the global consumption of all customers, as shown in Table 6 and in Fig. 7. This is a perfect indication of the type of customers subject to anomaly, who are rather few consumers.

Table 6
Amount of cleaned energy in GWh, distributed per cluster.

Cluster	Uncleaned	Cleaned	Rate
1	805.85	807.45	0.20%
2	349.35	340.32	2.59%
3	501.22	494.62	1.32%
4	665.81	655.26	1.59%
5	440.77	432.24	1.93%
6	293.50	294.74	0.42%
7	824.48	825.10	0.08%
8	346.30	346.02	0.08%
Global	4227.28	4195.75	0.75%

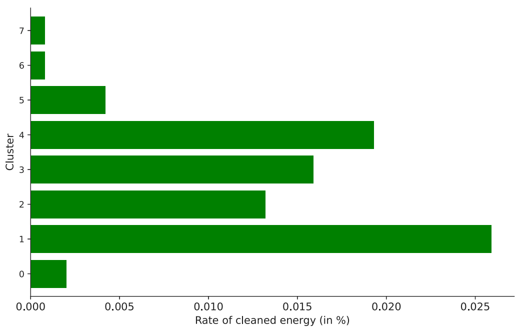

Fig. 7. Rate of cleaned energy, distributed per cluster.

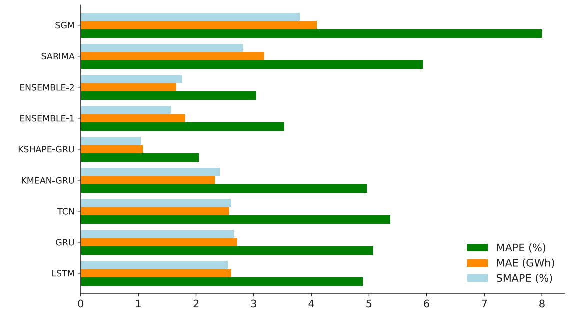

Fig. 8. Global MAPE, MAE and sMAPE errors for different methods.

5.3. Prediction result

This section discusses the results achieved through two different analyses: global results (Section 5.3.1) and individual results (Section 5.3.2). Global results describe predictions that aggregate the individual consumption of all consumers, while individual results refer to the prediction error measured for each individual separately. Notably, when making global predictions, any over- or under-prediction effects tend to cancel out, whereas such compensation does not occur in individual predictions.

5.3.1. Global result

The analysis by month of the K-Shape-GRU method from Tables 7–9 shows, overall, quite remarkable results. In particular, Table 7 provides the average monthly MAPE for different models over a period of one year. Firstly, we can see that the performance of the models varies across different months. For example, in January, the K-Shape-GRU model has the lowest MAPE (0.01%), while in August, it has one of the highest MAPE values (7.45%). This suggests that the performance of the models is affected by different factors in different months. It is important to note that the MAPE values for all models are generally higher during the summer months (June to August) and lower during the winter months (December to February). This could be due to seasonal changes in consumer behavior, market trends, or other external factors. Looking at the average MAPE values across all months, we can see that the K-Shape-GRU model performs the best with an average MAPE of 2.04%. This is followed by the SARIMA model (with an average MAPE of 5.93%) and the ENSEMBLE-1 and ENSEMBLE-2 models (both with an average MAPE of 3.53% and 3.04%, respectively), which had already reached remarkable results (Hadjout et al., 2022). This indicates that homogeneous clustering improves considerably the learning of the GRU model. The other interesting finding is that the comparison between K-Shape-GRU and K-Means-GRU indicates that K-Shape for our use case is more suitable for clustering of electricity consumption which is due to the nature of the BSD similarity measure.

Table 8 shows the average monthly MAE in GWh (gigawatt-hours) for different algorithms, including two proposed methods (K-Means-GRU and K-Shape-GRU), over a period of twelve months. From the table, we can see that the proposed K-Shape-GRU algorithm has the lowest MAE values for four out of twelve months (January, April, May, and November). The K-Means-GRU algorithm has the lowest MAE value for one month (September). However, for the remaining seven months, the performance of the proposed algorithms is not as good as some of the other methods. For instance, in July and August, both K-Shape-GRU and K-Means-GRU have relatively high MAE values compared to other algorithms. The table also shows that both LSTM and GRU algorithms have comparable performance, with an average MAE of 2.61 and 2.71 GWh, respectively. The TCN algorithm has a slightly lower average MAE of 2.57 GWh, while the SGM algorithm has the highest average MAE of 4.09 GWh. The two ensemble methods (ENSEMBLE-1 and ENSEMBLE-2) have relatively low MAE values, indicating that combining different algorithms can improve the prediction accuracy.

Table 9 shows the average monthly errors, measured by sMAPE, for various time series forecasting models. The results indicate that the proposed method, K-Shape-GRU, outperformed all other models in most months with an average sMAPE of 1.04%, followed by ENSEMBLE-1 and LSTM with an average sMAPE of 1.56% and 2.55%, respectively. In contrast, some models such as SARIMA and SGM consistently showed higher sMAPE errors across most months. Overall, the table suggests that the proposed method, K-Shape-GRU, can be a promising approach for accurately forecasting the given time series data. sMAPE is symmetric in the sense that it treats over- and under-predictions equally, and it is less sensitive to extreme values than MAPE. This effect is evident in the summer data, where the differences in sMAPE between the proposed method and other methods are significantly smaller.

Comparing these tables, we can see that there are some differences in the ranking of the models. For example, the LSTM model performs better in terms of MAE compared to the K-Shape-GRU model for some months, but it has a higher average MAPE value. This suggests that the LSTM model may be better at predicting overall trends in the data, while the K-Shape-GRU model may be better at capturing smaller fluctuations in the data. This effect is smoothed when sMAPE is assess, given the nature of the metric that takes into count the direction of the error.

Fig. 8 has been included in order to illustrate the superiority of our K-Shape-GRU approach compared to other known methods in the literature. Moreover, Fig. 9 is included to depict the comparison of predicted and actual values between the proposed approach and the different methods.

Tables 10 and 11 show the MAPE, MAE and sMAPE results by cluster of the two methods based on K-Means-GRU and K-Shape-GRU. It appears that these methods show a cluster with a bad result C6 for K-Means-GRU and C1 for K-Shape-GRU which is interesting for the in-depth analysis of the customers grouped in this kind of cluster.

To further analyze the results we investigated the influence of errors in each month on the overall errors. As shown in Fig. 10, the majority of errors stemmed from August (34%), followed by October (17%) and July (16%). These three months corresponded to a significant surge in

Table 7
Average MAPE monthly errors for both the individual, statistic and proposed method.

DATE	LSTM	GRU	TCN	ENSEMBLE-1	ENSEMBLE-2	SARIMA	SGM	K-Means-GRU	K-Shape-GRU
January	1.88%	5.60%	5.49%	0.66%	0.64%	1.04%	1.97%	5.75%	0.01%
February	0.42%	0.65%	7.88%	5.55%	1.85%	0.94%	10.64%	7.31%	2.35%
March	2.79%	1.99%	8.62%	0.72%	0.58%	7.33%	9.54%	6.13%	0.86%
April	6.18%	4.64%	5.42%	1.21%	2.59%	0.07%	4.41%	4.01%	0.56%
May	5.44%	2.49%	6.46%	6.46%	1.46%	5.44%	6.63%	5.13%	0.65%
June	4.99%	3.60%	6.37%	1.98%	1.50%	1.50%	7.08%	5.04%	0.09%
July	9.45%	12.64%	0.30%	2.65%	7.42%	2.56%	6.44%	1.19%	4.01%
August	13.50%	13.23%	1.79%	8.10%	10.16%	17.23%	15.52%	1.10%	7.45%
September	4.31%	3.05%	6.30%	2.47%	1.06%	7.35%	0.49%	6.60%	2.00%
October	8.43%	8.73%	0.14%	3.74%	6.18%	18.18%	15.28%	1.80%	4.04%
November	0.72%	0.64%	7.92%	4.25%	2.00%	5.92%	7.90%	7.58%	0.35%
December	0.61%	3.63%	7.71%	4.62%	1.06%	3.63%	10.01%	7.84%	2.09%
Average	4.89%	5.07%	5.37%	3.53%	3.04%	5.93%	7.99%	4.96%	2.04%

Table 8
Average MAE monthly errors for both the individual, statistic and proposed method (in GWh).

DATE	LSTM	GRU	TCN	ENSEMBLE-1	ENSEMBLE-2	SARIMA	SGM	K-Means-GRU	K-Shape-GRU
January	0.91	2.70	2.65	0.32	0.31	0.50	0.95	2.69	0.01
February	0.19	0.29	3.53	2.49	0.83	0.42	4.77	3.17	1.02
March	1.27	0.90	3.92	0.33	0.26	3.33	4.33	2.71	0.38
April	2.97	2.23	2.61	0.58	1.25	0.03	2.12	1.88	0.26
May	2.58	1.18	3.06	3.06	0.69	2.57	3.14	2.36	0.30
June	2.36	1.70	3.02	0.94	0.71	0.71	3.36	2.32	0.04
July	5.13	6.87	0.16	1.44	4.03	1.39	3.50	0.63	2.12
August	8.22	8.05	1.09	4.93	6.19	10.49	9.45	0.66	4.43
September	2.21	1.57	3.23	1.27	0.55	3.77	0.25	3.30	1.00
October	4.80	4.98	0.08	2.13	3.52	10.35	8.71	0.98	2.20
November	0.35	0.31	3.81	2.05	0.96	2.85	3.81	3.54	0.16
December	0.29	1.72	3.67	2.20	0.50	1.72	4.76	3.62	0.97
Average	2.61	2.71	2.57	1.81	1.65	3.18	4.09	2.32	1.07

Table 9
Average sMAPE monthly errors for both the individual, statistic and proposed method.

DATE	LSTM	GRU	TCN	ENSEMBLE-1	ENSEMBLE-2	SARIMA	SGM	K-Means-GRU	K-Shape-GRU
January	0.95%	2.88%	2.67%	0.32%	0.33%	0.51%	0.98%	2.80%	0.01%
February	0.21%	0.32%	3.79%	0.92%	2.70%	0.47%	5.05%	3.52%	1.16%
March	1.42%	1.00%	4.13%	0.29%	0.36%	3.54%	4.55%	2.97%	0.43%
April	3.19%	2.37%	2.64%	1.31%	0.61%	0.03%	2.16%	1.96%	0.28%
May	2.80%	1.26%	3.13%	0.74%	3.13%	2.65%	3.21%	2.50%	0.33%
June	2.56%	1.83%	3.09%	0.76%	0.98%	0.74%	3.42%	2.46%	0.04%
July	4.96%	6.75%	0.15%	3.85%	1.34%	1.27%	3.12%	0.59%	2.05%
August	7.24%	7.08%	0.90%	5.35%	4.22%	7.93%	7.20%	0.56%	3.87%
September	2.20%	1.55%	3.05%	0.53%	1.22%	3.55%	0.24%	3.19%	1.01%
October	4.40%	4.57%	0.07%	3.19%	1.91%	8.33%	7.10%	0.89%	2.06%
November	0.36%	0.32%	3.81%	0.99%	2.08%	2.87%	3.80%	3.65%	0.17%
December	0.31%	1.85%	3.71%	0.53%	2.26%	1.78%	4.77%	3.77%	1.04%
Average	2.55%	2.65%	2.60%	1.56%	1.76%	2.81%	3.80%	2.41%	1.04%

Table 10
Average MAPE, MAE and sMPAPE monthly errors for KMeans-GRU.

Cluster	MAPE	MAE	sMAPE
C1	2.93%	2.44	1.44%
C2	0.08%	0.03	0.04%
C3	1.80%	0.50	0.89%
C4	7.04%	5.32	3.65%
C5	4.51%	1.98	2.31%
C6	34.63%	22.16	14.76%
C7	2.36%	4.59	1.16%
C8	7.52%	4.18	3.62%

Table 11
Average MAPE, MAE and sMAPE monthly errors for KShape-GRU.

Cluster	MAPE	MAE	sMAPE
C1	18.66%	16.38	10.29%
C2	4.46%	5.36	2.18%
C3	1.00%	0.90	0.50%
C4	3.17%	3.45	1.61%
C5	3.61%	6.35	1.77%

electricity demand in 2019, thus contributing to the exceptional error rates observed in our analysis.

Experiments have shown that the best forecasting results are obtained for big food industry customers, particularly beverage companies. Such industries share similar energy consumption trends. For example, during the summer and Ramadan month, there is a high demand for energy due to the increased consumption of drinks. Noting that Ramadan month is the ninth month of the Islamic calendar but shifts ten days from the solar calendar each year. However, the impact of Ramadan month is the same in increasing the energy of companies of the same category. Conversely, during the winter, the consumption of beverages is lower, reducing the required energy for the same company category.

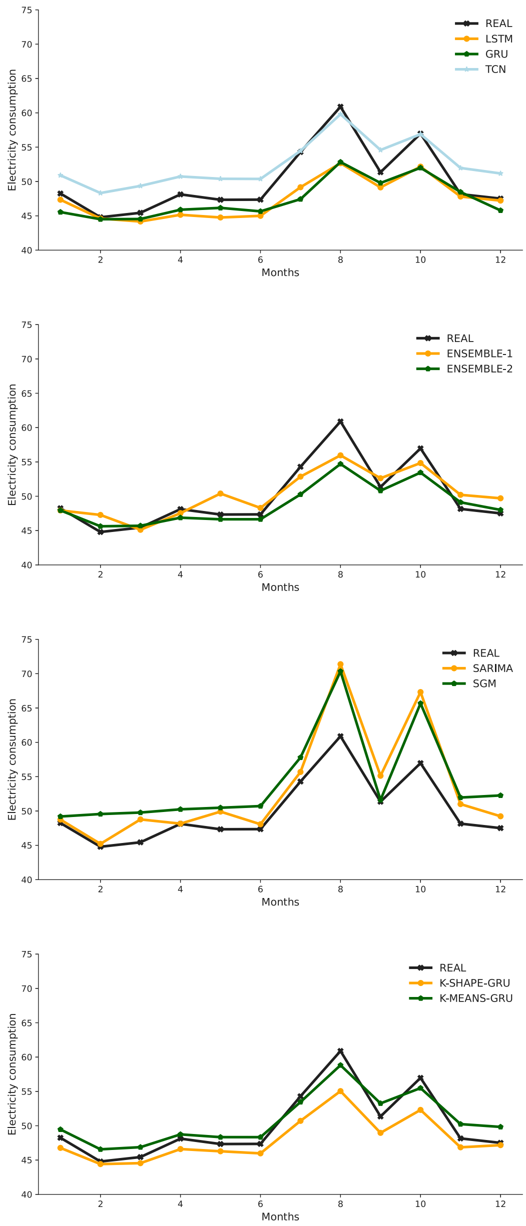

Fig. 9. Comparison of actual and predicted values for different models.

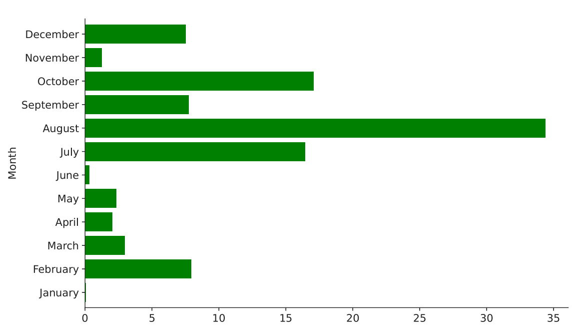

Fig. 10. Percentage impact month's errors on the aggregate results.

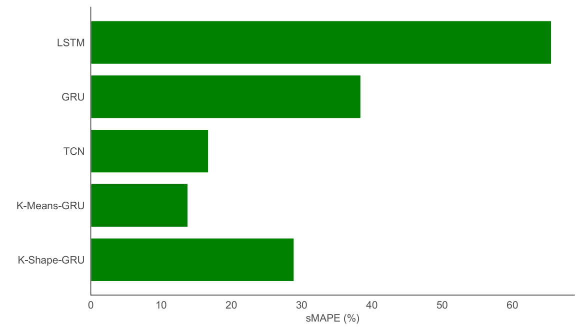

Fig. 11. Individual monthly sMAPE for the proposed approaches and individual deep learning models.

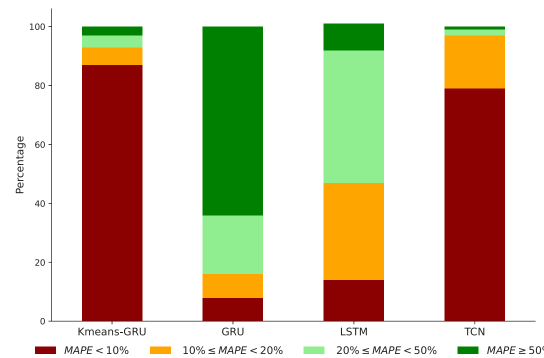

Fig. 12. Individual monthly MAPE for the proposed approach and deep learning methods LSTM, GRU and TCN.

5.3.2. Individual results

Individual consumption forecasting provides a powerful tool for the estimation of the consumption lump sum in case of customer's metering discrepancy and the calculation of the shortage to be invoiced in case of fraud.

Fig. 11 illustrates the sMAPE for the proposed method and the individual deep learning models. As discussed in Section 3.2.7, sMAPE is a particularly suitable metric to evaluate the impact on individual customer forecasts. The proposed K-Means-GRU method with 13.71%, method outperformed all other methods in terms of sMAPE error. TCN also achieved a noteworthy result, with a 16.75% sMAPE, indicating its potential for exploring individual electricity customer profiles in depth. K-Means-GRU achieved the third best result, thus confirming its potential for individual results as well.

Additionally, to evaluate the individual prediction per customer of our method we have analyzed the MAPE score (Moreno, Pol, S.Abad, & Blasco, 2013), given its easy interpretation. Thus, four score levels are considered:

1. Less than 10%, which is interpreted as a very good result.
2. Between 10 and 20%, which is interpreted as good results.
3. Between 20 and 50%, which is interpreted as bad results.
4. More than 50%, which is interpreted as a very bad result.

Therefore, we determine for each method the rate of scores per level and the result is given in Fig. 12. Its analysis shows the superiority of the clustering-based prediction methods by attaining 87% and 82% respectively for KMeans-GRU and KShape-GRU at a very good prediction level. However, the TCN method proves to be a competitor in client-based prediction.

6. Conclusions

In this research, we investigated prediction methods that combine deep learning, outlier detection, and clustering. We utilized the dataset of consumption by Bejaia economic consumers to assess the suggested approach in terms of MAPE, MAE and sMAPE. The proposed clustering method significantly enhances the prediction performance compared to other methods, according to the comparison results. In fact, the overall prediction result with the MAPE metric of 2.04% shows the effectiveness of clustering-based methods combined with anomaly detection and remediation methods when targeting accurate predictions. In addition, from an individual prediction point of view, the result obtained allowed us to design a practical solution that is integrated into the enterprise-level sales analysis system. TCN seems to be an interesting method to explore given its promising results for this kind of prediction.

This work does not include investigations and analysis into why specific months or clients resulted in significant errors, which could have enhanced the accuracy of our model. Additionally, upon analyzing the economic customer data, we find it compelling to explore the consumption patterns of residential or building-type customers, who constitute 90% of total customers but only 55% of total consumption. The challenge lies in accurately predicting the future consumption of each customer using a limited number of models. This research is crucial for reducing NTL and anticipating the impacts of decentralized production. Moreover, exogenous factors such as weather conditions, socio-economic factors, or geopolitical events could significantly affect model performance. Future research could explore how these exogenous factors can be incorporated into deep learning models to improve their accuracy and robustness.

CRediT authorship contribution statement

Dalil Hadjout: Conceptualization, Methodology, Software. **Abderrazak Sebaa:** Data curation, Writing – original draft. **José F. Torres:** Visualization, Investigation. **Francisco Martínez-Álvarez:** Supervision, Validation, Editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

The authors would like to thank the Spanish Ministry of Science and Innovation and the Junta de Andalucía, Spain for their support within the projects PID2020-117954RB-C21 and TED2021-131311B-C22, PY20-00870 and UPO-138516, respectively. Also, this work has been partially supported by the General Directorate of Scientific Research and Technological Development (DGRSDT, Algeria), under the PRFU project (ref: C00L07UN060120200003).

References

Bai, Y., Xie, J., Liu, C., Tao, Y., Zeng, B., & Li, C. (2021). Regression modeling for enterprise electricity consumption: A comparison of recurrent neural network and its variants. *International Journal of Electrical Power & Energy Systems*, *126*, Article 106612.

Bartos, M., Mullapudi, A., & Troutman, S. (2019). RRCF: Implementation of the Robust Random Cut Forest algorithm for anomaly detection on streams. *The Journal of Open Source Software*, *4*(35), 1336.

Bingchun, L., Chuanchuan, F., Arlene, B., & Quan, L. Y. (2017). Forecasting of Chinese primary energy consumption in 2021 with GRU artificial neural network. *Energies*, *10*(10), 1453.

Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. (2022). A review on outlier/anomaly detection in time series data. *ACM Computing Surveys*, *53*(id56), 1–33.

Box, G., & Jenkins, G. (2008). *Time series analysis: Forecasting and control*. John Wiley and Sons.

Braei, M., & Wagner, S. (2020). Anomaly detection in univariate time-series: A survey on the state-of-the-art. arXiv preprint arXiv:2004.00433.

Chikobvu, D., & Sigauke, C. (2012). Regression-SARIMA modelling of daily peak electricity demand in South Africa. *Journal of Energy in Southern Africa*, *23*(3), 23–30.

Cho, H., Kim, Y., Lee, E., Choi, D., Lee, Y., & Rhee, W. (2020). Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks. *IEEE Access*, *8*, 52588–52608.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *Proceedings of the NIPS 2014 workshop on deep learning* (pp. 1–9).

Divina, F., Torres, J. F., García-Torres, M., Martínez-Álvarez, F., & Troncoso, A. (2020). Hybridizing deep learning and neuroevolution: Application to the Spanish short-term electric energy consumption forecasting. *Applied Sciences*, *10*(16), 5487.

Fahiman, F., Erfani, S., Rajasegarar, S., Palaniswami, M., & Leckie, C. (2017). Improving load forecasting based on deep learning and K-Shape clustering. In *Proceedings of the international joint conference on neural networks* (pp. 4134–4141).

Gholizadeh, N., & Musilek, P. (2022). Federated learning with hyperparameter-based clustering for electrical load forecasting. *Internet of Things*, *17*, Article 100470.

Gupta, M., Gao, J., Aggarwal, C. C., & Han, J. (2013). Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *26*(9), 2250–2267.

Gutiérrez-Avilés, D., Fábregas, J. A., Tejedor, J., Martínez-Álvarez, F., Troncoso, A., Arcos, A., et al. (2018). SmartFD: A real big data application for electrical fraud detection. *Lecture Notes in Computer Science*, *10870*, 120–130.

Hadjout, D., Torres, J. F., Troncoso, A., Sebaa, A., & Martínez-Álvarez, F. (2022). Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market. *Energy*, *243*, Article 123060.

Jahangir, H., Tayarani, H., Gougheri, S., Golkar, M. A., Ahmadian, A., & Elkamel, A. (2020). Deep learning-based forecasting approach in smart grids with micro-clustering and bi-directional LSTM network. *IEEE Transactions on Industrial Electronics*.

Jeenanunta, C., Abeyrathna, K. D., Dilhani, M. S., Hnin, S. W., & Phyo, P. P. (2018). Time series outlier detection for short-term electricity load demand forecasting. *International Scientific Journal of Engineering and Technology*, *2*(1), 37–50.

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679.

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, *10*(1), 841–851.

Lara-Benítez, P., Carranza-García, M., Luna-Romera, J., & Riquelme, J. (2020). Temporal convolutional networks applied to energy-related time series forecasting. *Applied Sciences*, *10*(7), 2322.

Lara-Benítez, P., Carranza-García, M., & Riquelme, J. C. (2021). An experimental review on deep learning architectures for time series forecasting. *International Journal of Neural Systems*, *31*(3), Article 2130001.

Laurinec, P. (2018). Improving forecasting accuracy through the influence of time series representations and clustering. *Information Sciences and Technologies*, *10*(2), 6.

Lee, Y. S., & Tong, L. I. (2011). Forecasting energy consumption using a grey model improved by incorporating genetic programming. *Energy Conversion and Management*, *52*(1), 147–152.

Maronna, R. A. (2021). *Robust statistics: Theory and methods*. John Wiley and Sons.

Martínez-Álvarez, F., Troncoso, A., Asencio-Cortés, G., & Riquelme, J. C. (2015). A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, *8*(11), 13162–13193.

Martínez-Álvarez, F., Troncoso, A., Riquelme, C., & Aguilar-Ruiz, J. S. (2011). Discovery of motifs to forecast outlier occurrence in time series. *Pattern Recognition Letters*, *32*(12), 1652–1665.

Moreno, J. J. M., Pol, A. P., S.Abad, A., & Blasco, B. C. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, *25*(4), 500–506.

Nakyoung, K., Minkyung, K., & Kyun, C. J. (2018). LSTM based short-term electricity consumption forecast with daily load profile sequences. In *Proceedings of the IEEE global conference on consumer electronics* (pp. 136–137).

Paparrizos, J., & Gravano, L. (2015). K-Shape: Efficient and accurate clustering of time series. In *Proceedings of the ACM SIGMOD international conference on management of data* (pp. 1855–1870).

Pérez-Chacón, R., Luna-Romera, J. M., Troncoso, A., Martínez-Álvarez, F., & Riquelme, J. C. (2018). Big data analytics for discovering electricity consumption patterns in smart cities. *Energies*, *11*(3), 683.

Scitovski, R., Sabo, K., Martínez-Álvarez, F., & Ungar, S. (2021). *Cluster analysis and applications*. Springer.

Sepulveda, L. F. M., Diniz, P. S., Diniz, J. O. B., et al. (2021). Forecasting of individual electricity consumption using optimized gradient boosting regression with modified particle swarm optimization. *Engineering Applications of Artificial Intelligence*, *105*, Article 104440.

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., et al. (2020). Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, *21*(118), 1–6.

Torabi, M., Hashemi, S., Saybani, M. R., Shamshirband, S., & Mosavi, A. (2019). A hybrid clustering and classification technique for forecasting short-term energy consumption. *Environmental Progress and Sustainable Energy*, *38*(1), 66–76.

Torres, J. F., Galicia, A., Troncoso, A., & Martínez-Álvarez, F. (2018). A scalable approach based on deep learning for big data time series forecasting. *Integrated Computer-Aided Engineering*, *25*(4), 335–348.

Torres, J. F., Hadjout, D., Sebaa, A., Martínez-Álvarez, F., & Troncoso, A. (2021). Deep learning for time series forecasting: A survey. *Big Data*, *9*(1), 3–21.

Torres, J. F., Martínez-Álvarez, F., & Troncoso, A. (2022). A deep LSTM network for the Spanish electricity consumption forecasting. *Neural Computing and Applications*, *34*(13), 10533–10545.

Wu, J., Chen, X., Zhang, H., Xiong, L., Lei, H., & Deng, S. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), 26–40.

Yang, J., Ning, C., Deb, C., Zhang, F., Cheong, D., et al. (2017). K-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. *Energy and Buildings*, *146*, 27–37.