Greedy Clustering-based Monthly Electricity Consumption Forecasting Model

Yuqing Wang, Zhenghui Li, Fei Wang, Zhao Zhen
Department of Electrical Engineering
North China Electric Power University
Baoding 071003, China
feiwang@ncepu.edu.cn

Payman Dehghanian
Department of Electrical and Computer Engineering
George Washington University
Washington DC, USA
payman@gwu.edu

João P. S. Catalão
Faculty of Engineering
University of Porto and INESC TEC
Porto 4200-465, Portugal
catalao@fe.up.pt

Kangping Li
Department of Electrical Engineering
Tsinghua University
Beijing 100084, China
kangpingli@tsinghua.edu.cn

Mahmud Fotuhi-Firuzabad
Department of Electrical Engineering
Sharif University of Technology
Tehran 113658639, Iran
fotuhi@sharif.edu

2021 IEEE Industry Applications Society Annual Meeting (IAS) | 978-1-7281-6401-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/IAS48185.2021.9677327

***Abstract*—Accurate monthly electricity consumption forecasting is indispensable for electricity retailers to mitigate trading risks in the electricity market. Clustering-based forecasting method are commonly used to generate accurate monthly electricity consumption forecasting results. This paper focuses on the problem that the existing clustering-based monthly electricity consumption forecasting methods perform clustering and forecasting independently, causing that the joint optimization of two steps cannot be achieved. The reason for this situation is that the target of current clustering algorithms, maximizing individual similarity in a group, is not consistent with the final target of improving the forecasting accuracy. To solve the above problem, the greedy clustering-based monthly electricity consumption forecasting model (GCMECF) is proposed in this paper. Its clustering step takes improving the overall predictability as the optimization target, which is closely related to the forecasting target. In this way, with matching targets, the joint optimization of clustering and forecasting can be achieved. Meanwhile, the selection of the optimal number of clusters is decided based on the forecasting performance under multiple clustering scenarios. The case study verifies the effectiveness and superiority of the proposed method via a real-world dataset.**

Keywords*—*greedy clustering, clustering target, predictability, joint optimization, monthly electricity consumption forecasting

I. Introduction

Monthly electricity consumption forecasting, which concerns the estimation of the future electricity demand [1], can help the electricity retailers to master the customers' medium and long term electricity consumption. In this way, the electricity retailers can make reasonable purchasing decisions to trade in the wholesale market and mitigate the potential price risk caused by the huge fluctuations in the spot market [2-4]. Meanwhile, accurate monthly electricity consumption forecasting can also guide the planning department to reasonably arrange the mid-term operation and maintenance plan, reduce the cost of power supply and improve the reliability of the power grid [5].

In recent years, many methods have been applied on electricity consumption forecasting. From the perspective of forecasting algorithms, these methods can be classified into two categories: statistical methods and artificial intelligence (AI) methods [3]. The statistical methods mainly include autoregressive integrated moving average (ARIMA) [6], the grey model [7], etc. The AI methods mainly include artificial neural network (ANN) [8], support vector regression (SVR) [9] and recurrent neural network (RNN) [10], etc. From the perspective of spatial scale, it can be divided into regional forecasting which achieves the aggregated future electricity consumption of a region, and individual forecasting which achieves electricity consumption of lower aggregated level [11] (e.g., substation level, feeder level, and customer level). Generally, in the field of regional (aggregated) electricity consumption forecasting, there are three ways to obtain the forecasting results. The first way is to aggregate the electricity consumption data of each lower aggregated level object in this region to construct a regional level electricity consumption time series. Then, forecasting algorithms are performed on the regional time series to obtain the future regional electricity consumption [12], [13]. However, simply aggregating the lower level data cannot make full use of the spatially relevant information [14], [15]. Some papers investigate the second way to forecast the consumption data of each lower level object respectively, and then aggregate each forecasting results to get the final regional electricity consumption. However, due to the randomness and volatility of these individual series [16], [17], this approach cannot achieve an ideal forecasting results.

To effectively utilize relevant spatial information and reduce the negative effect on electricity consumption forecasting caused by randomness and volatility of individual series, some existing methods investigate the adoption of the third way, the clustering-based forecasting methods [11], [18-20]. In these methods, clustering algorithm is adopted to group all the individuals into several clusters first, then the forecasting algorithm is performed on each clusters separately to obtain forecasting results of each clusters, and eventually, the forecasting results of each clusters will be aggregated to obtain the final forecasting result. When the number of individual meets a certain threshold, the clustering-based forecasting methods can achieve an ideal forecasting results [21], [22]. However, in current clustering-based forecasting methods, the two steps of clustering and forecasting are executed separately and lack a direct linkage. The reason is that most of the current clustering algorithms, including K-means [23], the hierarchical clustering [24], and DBSCAN [25], take maximizing the intra-cluster similarity and minimizing the inter-cluster similarity under a predefined cluster number as clustering target but not the improvement of the final forecasting accuracy. In this way, forecasting and clustering are considered as two independent steps with weak correlation, and the improvement of the forecasting accuracy is restricted.

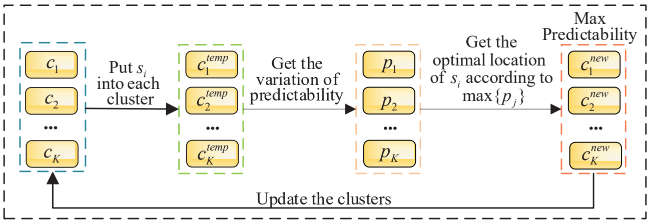

Fig. 1. The general process of greedy clustering for each object

To solve the above limitations, the greedy clustering-based electricity consumption forecasting model is proposed which directly couples the optimization target of clustering with the target of forecasting. As shown in Fig. 1, in the greedy clustering algorithm, firstly, K lower aggregated level objects are selected as the initial elements of the K clusters. Then, ordinally allocate the remaining objects s_i into the above K clusters according to certain criteria. The criteria is grouping the object into one of the clusters where it can achieve greater improvement of the predictability $\{p_j\}$. For each object, we only care about its optimal allocation in the current situation. In this way, the optimal partition of all individuals under a predefined K will be completed and the consistency of clustering target and forecasting target can be achieved.

In addition, the clustering algorithms were usually applied under a predefined cluster number K, so it is crucial to provide an appropriate cluster number to achieve the optimal partition of clusters [26]. In order to deal with this difficulty of choosing cluster number, some performance metrics, e.g., DBI, WCBCR, were used to find the appropriate number [27]. However, these metrics reflecting the intra-cluster compactness and inter-cluster separation are designed to optimize the conventional clustering algorithms, which also aims to maximize the similarity of each cluster but not the predictability. To conquer this problem, instead of determining the cluster number directly, a range of cluster number K which may include the optimal cluster number is set to generate multiple clustering scenarios. Then, the optimal cluster number is selected based on the monthly electricity forecasting accuracy of various clustering scenarios, which is performed on validation set.

Based on the above discussions, the contributions of this paper are as followed:

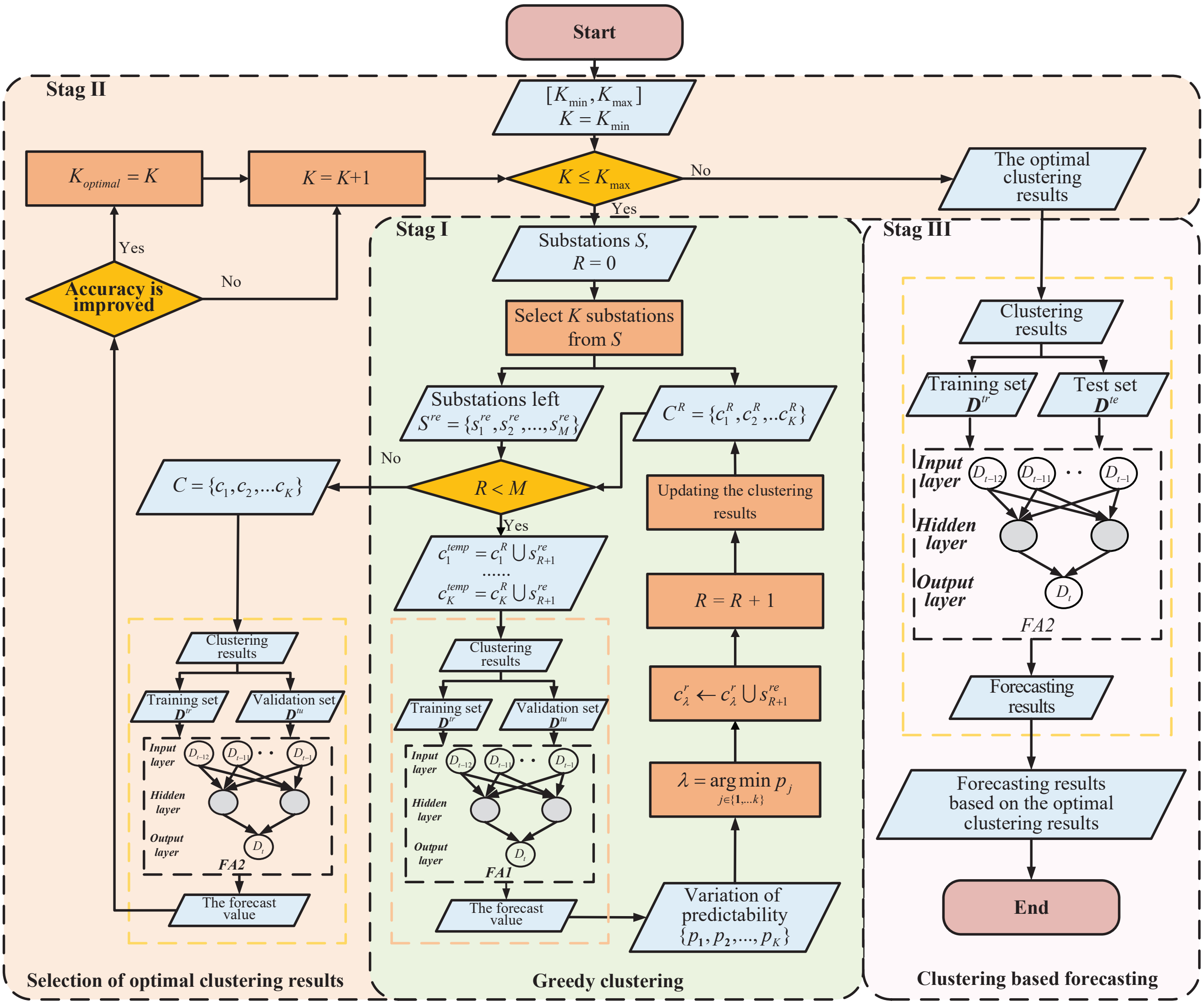

Fig. 2. The flowchart of greedy clustering-based forecasting model

1) The greedy clustering-based forecasting method directly links the clustering with forecasting by a novel greedy clustering algorithm which aims to maximize the predictability of each clusters. In this way, clustering and forecasting are modeled as a joint optimization problem to achieve the optimal allocation of each individuals under predefined cluster number, and improve the forecasting accuracy further.

2) The adaptive model to determine the optimal clustering number based on the forecasting accuracy under multiple clustering scenarios, which can avoid the difficulty to select the optimal cluster number without pre-knowledge or experience.

The rest of this paper is organized as follow: The greedy clustering-based forecasting model is proposed in Section II. The validation of this model on the real world data is present in Section III. Section IV draws the conclusion.

II. METHODOLOGY

This paper proposes a greedy clustering-based electricity consumption forecasting model, which align the clustering target with the improvement of forecasting accuracy. Furthermore, the optimal cluster number K can be automatically found. The whole process can be divided into three main stages, including the greedy clustering (stage I), the selection of optimal cluster number (stage II) and the clustering-based forecasting (stage III). The process is illustrated in Fig. 2.

In order to explain the method of this paper more clearly, the substations level data of a region are taken as an example. First, we presume that a region contains N substations $S=\{s_1,s_2,...,s_N\}$. D_t and $D_{m,t}$ are used to denote the electricity consumption data of total region S and the m-th substation s_m at time t. Then, the substation data $D_{m,t}$ is segmented into three parts: $D_{m,t}^{tr}$, $D_{m,t}^{va}$, $D_{m,t}^{te}$ for training, tuning and testing the proposed model.

In stage I, the greedy clustering algorithm is applied under a predefined cluster number, which maximizes the goodness of fit of the *Forecasting Algorithm 1(FA1)* on the data of each cluster to improve the overall predictability. Then, in stage II, the *Forecasting Algorithm 2 (FA2)* is applied based on different clustering scenarios to obtain the corresponding total electricity consumption forecasting results. After that, the optimal clustering results can be selected among them. The two stages both are performed on $\boldsymbol{D}^{tr}$ and $\boldsymbol{D}^{va}$.

Finally, in stage III, based on the optimal clustering results, the *FA2* will be employed on each cluster. Then, all of the forecasting results obtaining from each cluster will be aggregated to get the final regional monthly electricity consumption. This stage is conducted on $\boldsymbol{D}^{te}$.

A. The Proposed Greedy Clustering

The target of the greedy clustering algorithm is to put the substation $s_i \in S$ into the cluster $c_j \subset \{c_1,c_2,...c_K\}$, and the predictability of electricity consumption data of c_j can achieve the greater improvement. The predictability is mainly measured by goodness of fit of *FA1* on $\boldsymbol{D}^{va}$. With respect to the goodness of fit, many metrics and techniques have been presented in numerous publications. In this paper, the R^2 is used as the metric of it.

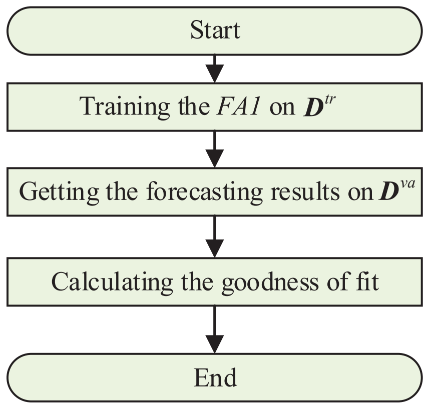

Fig. 3. The process of calculating the goodness of fit.

Algorithm 1: **Greedy clustering algorithm**

Input: The cluster number K, all substations $S=\{s_1,s_2,...,s_N\}$, electricity consumption data $\{\boldsymbol{D}_1,\boldsymbol{D}_2,...,\boldsymbol{D}_N\}$

Output: The clusters $C=\{c_1^R,c_2^R,...,c_K^R\}$

$S^{se}=\{s_1^{se},...,s_K^{se}\}$ ←Randomly select K substations from S;

$S^{re}=\{s_1^{re},...,s_M^{re}\}$ ←Remaining substations after selection;

$M=N-K, R=1$;

for $i\in\{1,...,K\}$ **do**

  $c_i^0 \leftarrow s_i$;

end for

for $R \le M$ **do**

  for c_j^{R-1} in $\mathbf{C}^{R-1}$ **do**

    $c_j^{temp}=c_j^{R-1}\bigcup s_R^{re}$;

    $\boldsymbol{D}_{c_j^{temp}}=\boldsymbol{D}_{c_j^0}+\boldsymbol{D}_i$;

    $p_j=(\phi_{c_j^{temp}}-\phi_{c_j^0})*|\boldsymbol{D}_{c_j^{temp}}|$;

  end for

  $\lambda=\arg\max_{j\in\{1,...k\}}\{p_j\}$;

  $c_\lambda^{R-1}\leftarrow c_\lambda^{R-1}\bigcup s_R^{re}$;

  $\mathbf{C}^R=\mathbf{C}^{R-1}$;

  $R=R+1$;

end for

The steps to calculate the goodness of fit are shown in Fig. 3. Take the consumption data of c_j for example. Firstly, the *FA1* is trained with the data from $\boldsymbol{D}^{tr}$. After obtaining the forecasting results on $\boldsymbol{D}^{va}$, the goodness of fit of the forecasting algorithm on c_j can be calculated according to Eq. (1) and (2).

$$\phi_{c_j}=\frac{1}{T}\sum_{t=1}^{T}\frac{(\hat{D}_{c_j,t}^{va}-\bar{D}_{c_j}^{va})^2}{(D_{c_j,t}^{va}-\bar{D}_{c_j}^{va})^2} \tag{1}$$

$$\bar{D}_{c_j}^{va} = \frac{1}{T}\sum_{t=1}^{T} D_{c_j,t}^{va} \tag{2}$$

Where ϕ_{c_i} denotes the goodness of fit of the forecasting algorithm on c_i; $\boldsymbol{D}_{c_j}$ represents the aggregated electricity consumption series of each elements from the *j*-th cluster c_j; $D_{c_j,t}^{va}$ and $\hat{D}_{c_j,t}^{va}$ denote the true value and forecasting value of $\boldsymbol{D}_{c_j}^{va}$ at time *t*; $\bar{D}_{c_j}^{va}$ denotes the average of $\boldsymbol{D}_{c_j}^{va}$, respectively; *T* represents the length of time series $\boldsymbol{D}_{c_j}^{va}$.

The whole procedures of greedy clustering are presented in **Algorithm 1** and the specific process of greedy clustering are as follows:

1) *Initialization:* K $(1 \le K \le N)$ substations $\mathbf{S}^{se} = \{s_1^{se}, s_2^{se}, \ldots s_K^{se}\}$ are randomly selected from the substations set S and the remaining substations $\mathbf{S}^{re} = \{s_1^{re}, s_2^{re}, \ldots, s_M^{re}\}$ is produced at the same time. $\mathbf{S}^{re}$ and $\mathbf{S}^{se}$ both satisfy the following constraints:

$$\begin{aligned} \mathbf{S}^{se} \bigcup \mathbf{S}^{re} &= \mathbf{S} \\ \mathbf{S}^{se} \bigcap \mathbf{S}^{re} &= \varnothing \end{aligned} \tag{3}$$

Making the S^{se} as the initial elements of *K* clusters $C^0 = \{c_1^0, c_2^0, \ldots, c_K^0\}$, i.e. $s_j^{se} \in c_j^0$, we can calculate the initial goodness of fit on each cluster to get $f^0 = [\phi_{c_1}^0, \phi_{c_2}^0, \ldots, \phi_{c_K}^0]$ using Eq. (1).

2) *Calculating the variation of predictability:* Arrange s_i^{re} from the remaining substations $\mathbf{S}^{re}$ into each cluster $c_j^0 \subset \{c_1^0, c_2^0, \ldots, c_K^0\}$ successively to generate $\mathbf{C}^{temp}$ as Eq. (4). And the electricity consumption data $\boldsymbol{D}_{c_j^{temp}}$ of c_j^{temp} can be obtained by Eq. (5)

$$\begin{aligned} \mathbf{C}^{temp} = \{ c_1^{temp} &= c_1^0 \bigcup s_i^{re}, \\ c_2^{temp} &= c_2^0 \bigcup s_i^{re}, \\ &\ldots \\ c_K^{temp} &= c_K^0 \bigcup s_i^{re} \} \end{aligned} \tag{4}$$

$$\boldsymbol{D}_{c_j^{temp}} = \boldsymbol{D}_{c_j^0} + \boldsymbol{D}_i \tag{5}$$

Then, we can get the new goodness of fit $f^1 = [\phi_{c_1}^1, \phi_{c_2}^1, \ldots, \phi_{c_K}^1]$ of cluster $c_j^{temp} \subset \mathbf{C}^{temp}$ using formula (1) again. The variation of goodness of fit before and after clustering s_i^{re} into c_j^0 could be a part of a description of predictability. Besides, it is worth noting that the cluster with larger magnitude of electricity consumption contributes more to the final forecasting result. Therefore, in the process of measuring the variation of the predictability, we take into account the impact of the magnitude of electricity consumption data of each cluster by multiplying the variation of goodness of fit with the total electricity consumption of the cluster c_j^{temp}. The variation of predictability $\{p_1, p_2, \ldots, p_K\}$ before and after clustering s_i^{re} into $c_j^0 \subset \mathbf{C}^0$ can be obtained by Eq. (6).

$$p_j = (\phi_{c_j^{temp}} - \phi_{c_j^0}) * |\boldsymbol{D}_{c_j^{temp}}| \tag{6}$$

where, $\phi_{c_j^0}$ and $\phi_{c_j^{temp}}$ represents the goodness of fit of consumption data of c_j^0 and c_j^{temp}; c_j^0 represents the *j*-th cluster of $\mathbf{C}^0$; $|\boldsymbol{D}_{c_j^{temp}}|$ represents the total electricity consumption of c_j^{temp}.

3) *Updating the clusters:* According to step 2, we can obtain the vector $\boldsymbol{P} = \{p_j\}$ to denote the variation of predictability. What we hope is that group s_i^{re} into the cluster c_λ^0 where s_i^{re} can achieve higher improvement of predictability. Therefore, after the comparsion of all the situations (put s_i^{re} into each cluster $\{c_1^0, \ldots, c_K^0\}$ to calculate variation of predictability) according to Eq. (7), we can obtain the corresponding cluster label λ of s_i^{re}, and then update the cluster c_λ^0 using Eq. (8) while other clusters of $\mathbf{C}^0$ keep the same. In this way, once clusters updating can be completed, the new cluster $\mathbf{C}^1 = \mathbf{C}^0$ can be obtain.

$$\lambda = \underset{j \in \{1, \ldots k\}}{\arg\max}\, \boldsymbol{P} \tag{7}$$

$$c_\lambda^0 \leftarrow c_\lambda^0 \bigcup s_i^{re} \tag{8}$$

where λ denotes the subscript corresponding to the minimum of P.

4) *Termination mechanism:* Repeating steps 2 and 3 $R = N - K$ times until there are no substations left in $\mathbf{S}^{re}$, and then the final clustering result $\{c_1^R, c_2^R, \ldots, c_K^R\}$, an optimal allocation of all substations under a predefined cluster number K can be obtained. It is considered to be the clustering result with the highest overall predictability under a predefined *K*.

B. The Selection of Optimal Cluster Number

The purpose of the greedy clustering is to achieve the maximum predictability of each cluster under a predefined cluster number *K*, while the selection of the optimal cluster number is to select the optimal cluster number which helps the forecasting model to minimize the Mean Absolute Percentage Error (MAPE) on $\boldsymbol{D}^{va}$.

To overcome the key challenge of predefining the optimal cluster number, this stage generates multiple clustering scenarios under different *K*. Based on these scenarios, *FA2* was used for each cluster to get *K* consumption forecasting results and then the *K* results are added together to obtain the forecast value of total electricity consumption. According to the Mean Absolute Percentage Error (MAPE) of the results under various clustering scenarios, $K_{optimal}$ can be obtain by Eqs. (9) and (10).

$$K_{optimal} = \underset{K\in[K_{\min},K_{\max}]}{\arg\min}\{\text{MAPE}\} \tag{9}$$

$$\text{MAPE} = \frac{100\%}{T}\sum_{t=1}^{T}\frac{|D_t^{va}-\hat{D}_t^{va}|}{D_t^{va}} \tag{10}$$

Where {MAPE} represents the set of all MAPE calculated under multiple forecasting scenarios; T represents the number of forecast values; D_t^{va} and $\hat{D}_t^{va}$ represents the total true value and total forecast value of electricity consumption at time t on $\boldsymbol{D}^{va}$.

In other words, aiming to determine the optimal cluster number, the greedy clustering is applied according to cluster number K which increases within this range $[K_{\min}, K_{\max}]$. Then, *FA2* is employed on $\boldsymbol{D}^{tr}$ and $\boldsymbol{D}^{va}$ to obtain the total consumption forecasting results based on each kind of clustering scenarios. As K increases, if the forecasting accuracy is improved (in other words, MAPE decreases), the optimal cluster number becomes the corresponding K. In this way, the optimal clustering results can be finally achieved. When this range is large enough, it will include the globally optimal K.

C. The Selection of Forecasting Algorithm

Forecasting algorithms are used in all three stages where the greedy clustering stage uses the *FA1*, the selection of optimal K stage, and the final forecasting stage use the *FA2*. Due to the forecasting model is not the main concern of this paper, the commonly used forecasting model, Extreme Learning Machine (ELM) [28], is applied on this greedy clustering based monthly electricity consumption forecasting model.

In this paper, the *FA1* and *FA2* both apply the same ELM, but it is worth noting that the forecasting algorithm applied in the first stage is only used to calculate the goodness of fit, so it definitely can be inconsistent with *FA2* but the input features and out features should stay same with *FA2*. As for the configuration, the hidden layer of ELM is set to 20 neurons, the kernel function is sigmoid, and no other optimizations are performing on it.

Considering the seasonal periodicity of electricity consumption, the input data is the electricity consumption data for the preceding 12 months of the month D_t to be forecasted.

$$\begin{gathered}\text{Input} = \{D_{t-12}, D_{t-11}, ..., D_{t-1}\}\\ \text{Output} = \{D_t\}\end{gathered} \tag{11}$$

Besides, to highlight the proposed method, only the historical electricity consumption data was used in this paper. Note that employing other factors (e.g. temperature) as input features can also help to improve the forecasting accuracy.

III. Case Study

A. Description of Dataset

The data used in this paper is collected from Ausgrid [29]. From all the substations operating continuously from May 1st of 2012 to Apr.30th of 2019, we selected 105 substations that have relatively complete data. After filling up the missing data through linear interpolation, the original data with 15-min sampling interval was aggregated into monthly data. Data from May 1st of 2012 to April 30th of 2017, from May 1st of 2017 to April 30th of 2018, and from May 1st of 2018 to April 30th of 2019 are used as training set, tuning (validation) set and testing set.

B. The Evaluation Metrics

In this paper, three commonly used metrics are adopted to quantify the forecasting performance of the proposed method, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE).

$$\text{MAE} = \frac{1}{T}\sum_{t=1}^{T}|D_t^{te}-\hat{D}_t^{te}| \tag{12}$$

$$\text{MAPE} = \frac{100\%}{T}\sum_{t=1}^{T}\frac{|D_t^{te}-\hat{D}_t^{te}|}{D_t^{te}} \tag{13}$$

$$\text{RMSE} = \frac{1}{T}\sqrt{\sum_{t=1}^{T}(D_t^{te}-\hat{D}_t^{te})^2} \tag{14}$$

Where T represents the number of forecast values; D_t^{te} and $\hat{D}_t^{te}$ respectively represent the true value and forecast value of electricity consumption at time t on $\boldsymbol{D}^{te}$.

C. The Configuration of Test

There are 105 substations in this dataset. For greedy clustering, if K is too large, it will increase the calculation burden; on the contrary, if K is too small, the obtained the clustering results may not meet the requirements of improving accuracy. According to the experience and considering both the forecasting accuracy and the complexity of model training, the range of K is set from 2 to 11. It is noticed that this range may not be large enough to include the globally optimal cluster number, but it is large enough to show the clustering process.

To verify the superiority of the proposed method, three other monthly electricity consumption forecasting methods are set as comparisons:

M1: The proposed greedy clustering-based monthly electricity consumption forecasting method.

M2: K-means based monthly electricity consumption forecasting method, i.e. replace the proposed greedy clustering method with K-means.

M3: Forecast after aggregated, i.e. the electricity consumption data of 105 substations are accumulated to construct the regional electricity consumption time series, and then ELM is used to forecast the accumulated time series directly.

TABLE I. ERROR COMPARISON OF DIFFERENT METHODS

Metrics	M1	M2	M3	M4
MAE/MWh	**45721.70**	57551.64	76960.82	226944082
MAPE/%	**3.30**	4.19	4.84	14.30
RMSE/MWh	**62973.07**	68230.40	76960.82	226944.82

M4: Forecast before aggregated, i.e. the electricity consumption data of all substations are forecasted separately by ELM, and then the 105 forecasting results are aggregated to get the final results.

Table I shows the forecasting metrics of these four methods, while the best metrics are bolded. Obviously, according to these metrics, comparing with the forecasting methods without clustering optimization (M3, M4), the clustering-based forecasting methods (M1, M2) achieves better forecasting results, which is well known. Among them, the M1 achieves the best MAPE at 3.30%, while the M4 has the worst MAPE at 14.30%. It can also be seen that the proposed M1 model achieves lower RMSE, MAE, and MAPE than M2, which clustering target is inconsistent with the forecasting target.

Fig.4 shows the curve between the real values and the forecast values of the four methods. It can be seen that the M1 also has a good fitting effect.

The forecast results of all methods are recorded in Table II, and the optimal results of each month are in grey. It clearly shows that the proposed M1 method achieves the best forecast results for most months.

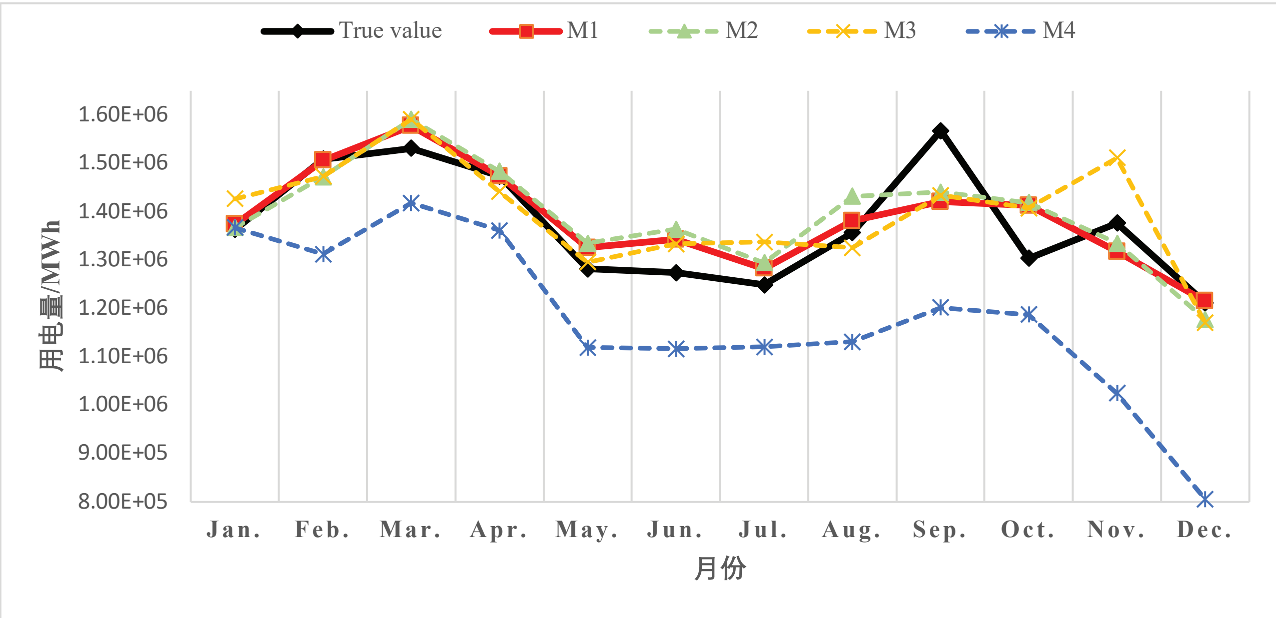

Fig. 4. Forecasting result curves of different methods

TABLE II. THE FORECASTING RESULTS OF DIFFERENT METHODS

Month	True value/MWh	Forecast value/MWh			
		GCMECF	*KMECF*	*FAA*	*FBA*
Jan.	1363355.10	1373771.44	1366895.59	1426803.23	1367405.87
Feb.	1508640.66	1505798.08	1472059.42	1473205.76	1312456.51
Mar.	1531246.07	1577306.12	1590790.33	1591456.02	1418131.13
Apr.	1473714.19	1473621.34	1483263.07	1440980.88	1361458.59
May.	1281689.57	1325807.08	1334473.09	1295797.31	1119057.17
Jun.	1274009.60	1342828.34	1363985.81	1333729.76	1116763.46
Jul.	1248956.83	1282358.08	1294413.55	1337497.11	1120441.15
Aug.	1356315.36	1381294.00	1431986.29	1325209.08	1130942.53
Sep.	1567507.58	1420791.25	1440533.81	1433768.04	1202105.22
Oct.	1303756.19	1413003.76	1418604.89	1407527.14	1187200.77
Nov.	1376441.71	1317650.39	1335018.02	1511111.00	1024986.70
Dec.	1211899.12	1215076.29	1177627.81	1170250.55	805651.67

TABLE III. ERROR COMPARISON OF DIFFERENT MEHODS

Metrics	M1	G1	G2	G3	G4
MAE/MWh	45721.70	52395.67	49200.50	47000.20	43960.46
MAPE/%	3.30	3.79	3.51	3.40	3.19
RMSE/MWh	62973.07	64778.59	63832.63	59057.08	54927.85

D. The Comparison of Using Different Forecasting Algorithm to Calculate the Goodness of Fit

In the previous section, both the *FA1* and *FA2* of the proposed method use the same ELM algorithm. It is worth noting that the *FA1* is only used to calculate the goodness of fit, so it does not have to stay the same with *FA2*.

In this selection, aiming to ensure the effectiveness of this method when using different forecasting algorithms as *FA1* and keeping *FA2* the same.Four kind of greedy clustering based forecasting models are designed as follow:

G1: Linear Regression (LR) with default parameters is used as *FA1*.

G2: Support Vector Regression (SVR) with a linear kernel is used as *FA1*.

G3: ELM which has 50 hidden layers of neurons is used as *FA1*.

G4: ELM which has 100 hidden layers of neurons is used as *FA1*.

The metrics of forecasting performance are recorded in TABLE III. Comparing with the forecasting results in the previous section, these four greedy clustering based forecasting models all achieve better forecasting accuracy than those of conventional methods. However, the results also indicate that the selection of the *FA1* may influence the final forecasting accuracy when *FA2* keeps the same. Only based on the forecasting performance metrics of these models, we can infer that when *FA1* and *FA2* are the same types, the M1 model can achieve better forecasting results. But the truth of the problem and its reasons need to be further explored in future research.

IV. Conclusion

This paper proposes a novel greedy clustering-based forecasting method for monthly electricity consumption forecasting. The greedy clustering takes the improvement of the overall predictability as to the fundamental clustering target which matches well with the forecasting target of improving accuracy. In this clustering step, the goodness of fit and the magnitude of electricity consumption are used to measure the variation of predictability. Meanwhile, to select the optimal cluster number for the proposed greedy clustering, a forecasting performance based automatically selecting method was applied in this GCMECF model. The real-world data from Ausgrid demonstrates the superior performance of the proposed greedy clustering-based forecasting method when compared with the commonly used K-means based forecasting method which considers clustering and forecasting separately and two other conventional methods. However, it's worth noting that this greedy clustering algorithm may be caught into the locally optimal situation, which may have some adverse effects on the forecasting accuracy. We will try to solve this problem in future research.

Acknowledgment

This work was supported by the National Key R&D Program of China (2018YFE0122200). This work was supported by the National Key R&D Program of China (2018YFE0122200). Also, J.P.S. Catalão acknowledges the support by FEDER through COMPETE 2020 and FCT, under POCI-01-0145-FEDER-029803 (02/SAICT/2017).

References

[1] M. Hu, J. Duan, Z. Hou, P. Wang, S. Fang, P. Hou, C. Fu, "Monthly electricity forecast based on electricity consumption characteristics analysis and multiple effect factors," *2019 IEEE 8th International Conference on Advanced Power System Automation and Protection (APAP)*, Xi'an, China, 2019, pp. 1814-1818.

[2] F. Wang, X. Ge, K. Li, and Z. Mi, "Day-ahead market optimal bidding strategy and quantitative compensation mechanism design for load aggregator engaging demand response," *IEEE Trans. Ind. Appl.*, vol. 55, no. 6, pp. 5564–5573, 2019.

[3] Z. Li, K. Li, F. Wang, Z. Xuan, Z. Mi, W. Li, P. Dehghanian, and M. Fotuhi-Firuzabad, "Monthly Electricity Consumption Forecasting: A Step-Reduction Strategy and Autoencoder Neural Network," *IEEE Ind. Appl. Mag.*, vol. 27, no. 2, pp. 90–102, 2021

[4] X. Lu, K. Li, H. Xu, F. Wang, Z. Zhou, and Y. Zhang, "Fundamentals and business model for resource aggregator of demand response in electricity markets," *Energy*, vol. 204, no. 117885, 2020.

[5] W. Jiang, X. Wu, Y. Gong, W. Yu, and X. Zhong, "Holt–winters smoothing enhanced by fruit fly optimization algorithm to forecast monthly electricity consumption," *Energy*, vol. 193, p. 116779, 2020.

[6] C. Yuan, S. Liu, and Z. Fang, "Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM(1,1) model," *Energy*, vol. 100, pp. 384–390, 2016.

[7] L. Wu, X. Gao, Y. Xiao, Y. Yang, and X. Chen, "Using a novel multi-variable grey model to forecast the electricity consumption of Shandong province in China," *Energy*, vol. 157, no. 2018, pp. 327–335, 2018.

[8] F. Wang, Y. Yu, Z. Zhang, J. Li, Z. Zhen, and K. Li, "Wavelet decomposition and convolutional LSTM networks based improved deep learning model for solar irradiance forecasting," *Appl. Sci.*, vol. 8, no. 8, pp. 1–29, 2018.

[9] Z. Hu, Y. Bao, R. Chiong, and T. Xiong, "Mid-term interval load forecasting using multi-output support vector regression with a memetic algorithm for feature selection," *Energy*, vol. 84, pp. 419–431, 2015.

[10] Z. Li, K. Li, F. Wang, Z. Mi, W. Li, and P. Dehghanian, "Auto-encoder neural network-based monthly electricity consumption forecasting method using hourly data," 2020 IEEE/IAS 56th Industrial and Commercial Power Systems Technical Conference *(I&CPS)*, Las Vegas, NV, USA, 2020, pp. 1-8.

[11] K. Li, F. Wang, Z. Mi, M. Fotuhi-Firuzabad, N. Duić, and T. Wang, "Capacity and output power estimation approach of individual behind-the-meter distributed photovoltaic system for demand response baseline estimation," *Appl. Energy*, vol. 253, p. 113595, 2019.

[12] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecast.*, vol. 32, no. 3, pp. 914–938, 2016.

[13] T. Hong, J. Xie, and J. Black, "Global energy forecasting competition 2017: Hierarchical probabilistic load forecasting," *Int. J. Forecast.*, vol. 35, no. 4, pp. 1389–1399, 2019.

[14] C. Zhang and R. Li, "A novel closed-loop clustering algorithm for hierarchical load forecasting," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 432–441, 2021.

[15] F. L. Quilumba, W. J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 911–918, 2015.

[16] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Trans. Smart Grid*, vol. 33, no. 1, pp. 1087-1088, 2018.

[17] L. Li, C. J. Meinrenken, V. Modi, and P. J. Culligan, "Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features," *Appl. Energy*, vol. 287, p. 116509, 2021.

[18] Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An ensemble forecasting method for the aggregated load with subprofiles," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3906–3908, 2018.

[19] X. Fu, X. J. Zeng, P. Feng, and X. Cai, "Clustering-based short-term load forecasting for residential electricity under the increasing-block pricing tariffs in China," *Energy*, vol. 165, pp. 76–89, 2018.

[20] C. H. Jin, G. Pok, Y. Lee, H. W. Park, K. D. Kim, U. Yun, and K. H. Ryu, "A SOM clustering pattern sequence-based next symbol

prediction method for day-ahead direct electricity load and price forecasting," *Energy Convers. Manag.*, vol. 90, pp. 84–92, 2015.

[21] F. Wang, B. Xiang, K. Li, X. Ge, H. Lu, J. Lai, P. Dehghanian, "Smart households' aggregated capacity forecasting for load aggregators under incentive-based demand response programs," *IEEE Trans. Ind. Appl*., vol. 56, no. 2, pp. 1086-1097, 2020.

[22] T. K. Wijaya, M. Vasirani, S. Humeau, and K. Aberer. Cluster-based aggregate forecasting for residential electricity demand using smart meter data[C]. *2015 IEEE International Conference on Big Data (Big Data)*, Santa Clara, CA, USA, 29 Oct.-1 Nov. 2015.

[23] X. Wang, W. J. Lee, H. Huang, R. L. Szabados, D. Y. Wang, and P. V. Olinda, "Factors that impact the accuracy of clustering-based Load forecasting," *IEEE Trans. Ind. Appl.*, vol. 52, no. 5, pp. 3625–3630, 2016.

[24] A. M. Alonso, F. J. Nogales, and C. Ruiz, "Hierarchical clustering for smart meter electricity loads based on quantile autocovariances," *IEEE Trans. Smart Grid*, vol. 11, no. 5, pp. 4522–4530, 2019.

[25] K. Li, X. Cao, X. Ge, F. Wang, X. Lu, M. Shi, R. Yin, Z. Mi, and S. Chang, "Meta-heuristic optimization-based two-stage considering intra-cluster compactness and inter-cluster separation," *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 3375–3384, 2020.

[26] N. B. Nikhare and P. S. Prasad, "A review on inter-cluster and intra-cluster similarity using bisected fuzzy C-mean technique via outward statistical testing," pp. 806–808, 2018.

[27] F. Wang, K. Li, C. Liu, Z. Mi, M. Shafie-Khah, and J. P. S. Catalao, "Synchronous pattern matching principle-based residential demand response baseline estimation: Mechanism analysis and approach description," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6972–6985, 2018.

[28] G. Huang, "What are Extreme Learning Machines ? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle," *Cognit. Comput.*, vol. 7, no. 3, pp. 263–278, 2015.

[29] Ausgrid, "Distribution zone substation information data to share," https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Distribution-zone-substation-data, accessed, August 31, 2020