A Novel Closed-Loop Clustering Algorithm for Hierarchical Load Forecasting

Chi Zhang, *Graduate Student Member, IEEE*, and Ran Li, *Member, IEEE*

***Abstract*—Hierarchical load forecasting (HLF) is an approach to generate forecasts for hierarchical load time series. The performance of HLF can be improved by optimizing the *forecasting model* and the *hierarchical structure*. Previous studies mostly focus on the forecasting model while the hierarchical structure is usually formed by clustering of natural attributes like geography, customer type, or the similarities between load profiles. A major limitation of these natural hierarchical structures is the mismatched objectives between clustering and forecasting. Clustering aims to minimize the dissimilarity among customers of a group while forecasting aims to minimize their forecasting errors. The two independent optimizations could limit the overall forecasting performance. Hence, this paper attempts to integrate the hierarchical structure and the forecasting model by a novel closed-loop clustering (CLC) algorithm. It links the objectives of forecasting and clustering by a feedback mechanism to return the goodness-of-fit as the criterion for the clustering. In this way, the hierarchical structure is enhanced by re-assigning the cluster membership and the parameters of the forecasting models are updated iteratively. The method is comparatively assessed with existing HLF methods. Using the same forecasting model, the proposed hierarchical structure outperforms the bottom-up structure by 52.20%, ensemble-based structure by 26.89%, load-profile structures by 19.90%, respectively.**

***Index Terms*—Load forecasting, hierarchical forecasting, smart meter, clustering, big data.**

I. Introduction

A. Problem Formulation

HIERARCHICAL forecasting is an approach to build models and generate forecasts for time series with a hierarchical structure. For a better understanding, Fig. 1 represents a 2-level hierarchical structure. Each node represents one time series and each row represents one level. Each level (except the bottom one) is the aggregation of time series from its lower levels.

The objective of hierarchical forecasting is to optimize the forecasting performance at a given level or all levels in the hierarchy. One application of hierarchical forecasting in power system is called hierarchical load forecasting (HLF), where we usually focus on the most aggregated level, such as regional or substation levels, utilizing data at the bottom level, e.g., smart meters [1], [2].

Manuscript received January 27, 2020; revised May 2, 2020 and June 11, 2020; accepted August 2, 2020. Date of publication August 7, 2020; date of current version December 21, 2020. Paper no. TSG-00124-2020. *(Corresponding author: Ran Li.)*

The authors are with the Department of Electronic and Electrical Engineering, University of Bath, Bath, BA2 7AY, U.K. (e-mail: cz382@bath.ac.uk; r.li2@bath.ac.uk).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2020.3015000

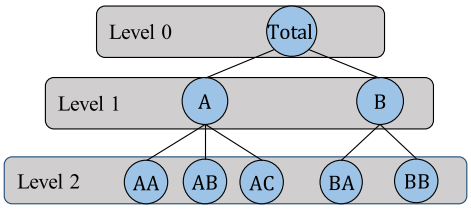


Fig. 1. A 2-level hierarchical structure.

Let $\mathbf{Y} = \{(y_{m,t})\} \in \mathbb{R}_{+}^{M\times T}$ be a non-negative matrix formed by the smart metering data of $M \in \mathbf{Z}^{+}$ customers in a region. Each customer $m \in M$ has T observations of electricity consumption as a time series $\boldsymbol{y}_m = (y_{m,1}, y_{m,2}, \ldots, y_{m,T})$. The regional electricity consumption over the period $t \in (1, 2, \ldots, T)$ can be expressed as the sum of $K(1 \le K \le M)$ groups of customers, denoted as $\boldsymbol{S}$.

$$\begin{aligned} \boldsymbol{S} &= \sum_{k=1}^{K} \sum_{m\in C_k} \boldsymbol{y}_m, \\ \boldsymbol{C} &= \{C_1, C_2, \ldots, C_k, \ldots, C_K\} \end{aligned} \tag{1}$$

Where $\boldsymbol{C}$ is the hierarchical structure (i.e., group set) and C_k is the k-th group of customers.

The objective of HLF is to minimize the forecasting error over the forecasting period at the regional level. For demonstration, we take mean absolute error (MAE) as the forecasting error metric denoted in (2).

$$\text{MAE} = \min \sum_{h=1}^{H} \left| \hat{S}_{T+h} \middle| F(\boldsymbol{\varphi}) - S_{T+h} \right| \tag{2}$$

Where $\hat{S}_{T+h}$ is the h-step-ahead forecast for the regional electricity consumption; S_{T+h} is the corresponding actual value of $\hat{S}_{T+h}$.

The performance of HLF can be advanced either by optimizing the forecasting model F or the hierarchical structure $\boldsymbol{C}$. In HLF, the optimization of forecasting models has been widely investigated and discussed [3]–[7]; strategies regarding optimizing hierarchical structure have been reported, including three approaches: top-down, bottom-up and middle-out [8].

B. Existing Literature and Limitations

1) Top-Down Approach: Fig. 2 is the general process of the top-down approach. The future regional electricity

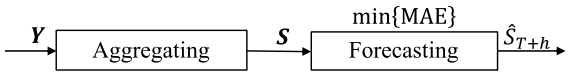


Fig. 2. General process of the top-down approach.

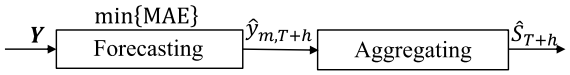


Fig. 3. General process of the bottom-up approach.

consumption is derived by directly fitting $\boldsymbol{S}$ as in (3) [4], [6]:

$$\hat{S}_{T+h} = F^{S}(\boldsymbol{S}; \boldsymbol{\varphi}) \tag{3}$$

Where F^S is the forecasting model fitted by $\boldsymbol{S}$; $\boldsymbol{\varphi}$ is the set of parameters or hyper-parameters of F depending on whether F is a parametric or nonparametric model.

2) Bottom-Up Approach: Fig. 3 is the general process of the bottom-up approach. The approach utilizes time series at the smart-meter level $\boldsymbol{Y}$ [9]–[11]. It first generates independent forecasts $\hat{y}_{m,T+h}$ of each individual customer m by fitting $\boldsymbol{y}_m$ as in (4). Then the regional electricity consumption $\hat{S}_{T+h}$ is obtained through the aggregation of the individual's electricity consumption, as in (5):

$$\hat{y}_{m,T+h} = F^{m}\left(\boldsymbol{y}_m; \boldsymbol{\varphi}^{m}\right), m \in M \tag{4}$$

$$\hat{S}_{T+h} = \sum_{m=1}^{M} \hat{y}_{m,T+h} \tag{5}$$

Where, for customer m, $\hat{y}_{m,T+h}$ is the h-step-ahead forecast; F^m is the forecasting model and $\boldsymbol{\varphi}^m$ is the parameters or hyper-parameters for F^m.

3) Middle-Out Approach: As for the middle-out approach, it starts at an intermediate level in the hierarchy and the regional forecasts are generated using the bottom-up strategy by aggregating the "middle-level" forecasts. Two categories of middle-out approaches are identified in the literature:

a) Clustering-based methods: Clustering is a type of unsupervised learning method to divide sample points into a number of groups so that points within the group shares similar properties or attributes. Considering approaches to identify the points of a cluster, clustering methods can be classified into density-based, centroid-based, distribution-based methods [12]. In HLF, clustering-based methods attempt to reduce the variance by grouping similar customers prior to forecasting [13]–[15]. Considering partitioning M customers into $K(1 \le K \le m)$ sets $\boldsymbol{C} = \{C_1, C_2, \ldots, C_k, \ldots, C_K\}$, the similarity measurement can be expressed as follows:

$$\arg\min_{\boldsymbol{C}} \sum_{k=1}^{K} \sum_{y_m \in C_k} \left[d(\boldsymbol{y}_m)\right], \ m \in M \tag{6}$$

Where $d(\boldsymbol{y}_m)$ is a distance-based metric to measure the within-cluster similarities.

With the established cluster C_k, forecasting model F^{C_k} with its parameters $\boldsymbol{\varphi}^{C_k}$ can be derived by fitting the sum series of the cluster as in (7). The regional forecasts $\hat{S}_{T+h}$ are generated through aggregating the forecasts from all clusters $\hat{S}^{C_k}_{T+h}$ as in (8).

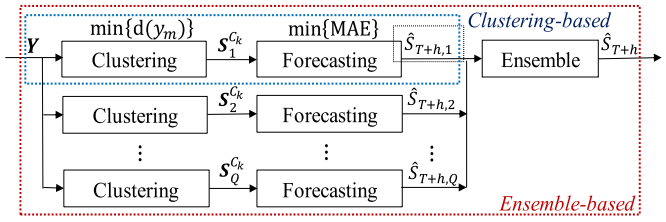


Fig. 4. General process of the middle-out approach.

$$\hat{S}^{C_k}_{T+h} = F^{C_k}\left(\sum_{\boldsymbol{y}_m \in C_k} \boldsymbol{y}_m; \boldsymbol{\varphi}^{C_k}\right) \tag{7}$$

$$\hat{S}_{T+h} = \sum_{k=1}^{K} \hat{S}^{C_k}_{T+h} \tag{8}$$

b) Ensemble-based methods: The key challenge of clustering-based methods is how to identify the optimal number of clusters K. The ensemble-based methods get around this challenge by generating multiple forecasts through varying K to increase the coverage on the optimal number of clusters. The final results are derived as a weighted average of all forecasts. The weight vector $\boldsymbol{\omega}$ is optimized as follows:

$$\begin{aligned} \boldsymbol{\omega} = \arg\min_{\boldsymbol{\omega}} & \sum_{h=1}^{H} \frac{\left|\hat{S}_{T+h} - S_{T+h}\right|}{S_{T+h}} \\ s.t. \ \ & \hat{S}_{T+h} = \sum_{q=1}^{Q} \omega_q \hat{S}_{T+h,q} \\ & \sum_{q=1}^{Q} \omega_q = 1, \omega_q \ge 0 \end{aligned} \tag{9}$$

Where $\boldsymbol{\omega} = (\omega_1, \omega_2, \ldots, \omega_q, \ldots, \omega_Q)$ is the vector of weight coefficients for Q forecasts; Q is the total number of forecasts; q is the q-th forecast and the number of clusters K varies for each forecast; ω_q is the weight coefficient for the q-th forecast; $\hat{S}_{T+h,q}$ is the h-step-ahead regional forecasts for the q-th forecast.

Fig. 4 describes the general process of the middle-out approach. The red box is the schematic diagram for ensemble-based methods, which can be viewed as an extension of the clustering-based methods as depicted in the blue box. Each block represents one stage of the method with its objective function above. Clustering-based methods include two independent stages as clustering and forecasting with different objective functions. Ensemble-based methods repeat the process of clustering-based methods for multiple times with different values of K and the final results are taken as their weighted average. Both methods are open-loop design.

It is noted that the top-down and bottom-up approaches can be considered as two special cases of clustering-based methods as $K = 1$ and $K = M$, respectively. In this way, the optimization strategy in terms of hierarchical structure can be classified into two categories as clustering-based and

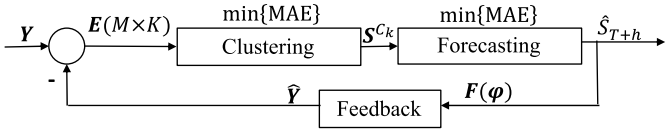


Fig. 5. General process of the proposed CLC method.

ensemble-based methods. As shown in Fig. 4, a major drawback of the clustering-based methods is the mismatch of the objectives between clustering and forecasting stages. The clustering stage aims to minimize the within-cluster distance (e.g., $\min\{d(y_m\})$ while the forecasting stage aims to minimize the forecasting error (e.g., $\min\{MAE\}$). The ensemble-based methods use the weighted average of multiple forecasts in hope to offset the impact but do not solve the fundamental problem. To conclude, the existing methods leave two problems unsolved:

1) For a given K, the optimal allocation of elements to clusters $\boldsymbol{C} = \{C_1, C_2, \ldots, C_k, \ldots, C_K\}$ are not determined under the criterion to advance the forecasting performance; In our case, the problem is converted to the determination of the optimal cluster membership vector $\boldsymbol{N} = \{n_1, n_2, \ldots, n_m, \ldots, n_M\}$, where $n_m \in \{1, 2, 3, \ldots, K\}$ is the cluster membership for customer m.
2) The challenge of the existing methods is the identification of the optimal number of clusters K given $\boldsymbol{Y}$. The searching for optimal K could be difficult without pre-knowledge or experience.

C. Contributions

The key contributions of this paper are as follows:

1) The paper proposes a closed-loop clustering (CLC) method to align the overall objective (e.g., $\min\{MAE\}$). As shown in Fig. 5, it is implemented through an extra feedback mechanism, which links the forecasting stage with the clustering stage. It returns the signal of model fitness and employs it as the clustering criterion to update the cluster membership $\boldsymbol{N}$. The customer portfolio, i.e., the elements within the cluster, is gradually improved by re-assigning load patterns to the established clusters throughout the iterative process for a given K.
2) The proposed CLC method can automatically search for the value of optimal K by a 'trim and merge' strategy, which leads to the optimal forecasts for the total electricity consumption $\hat{S}_{T+h}$.

The rest of the paper is organized as follows: Section II proposes the CLC method. Section III introduces the setup of the experiments on both simulated and real datasets. Followed by that, Section IV is the demonstration of the results, through comparative analysis with the existing HLF methods. Conclusions and potential areas of model improvement are drawn in Section V.

II. Proposed Methodology

This paper proposes an iterative closed-loop system to align the overall optimization objective and capture the ability to automatically find the optimal K. The detailed process is as follows.

A. Closed-Loop Clustering Method

1) Initialization: The method starts with a random assignment or clustering method such as K-means to partition M customers into $K(1 \le K \le M)$ sets

$$\begin{aligned} \boldsymbol{C}(0) &= \{C_1(0), C_2(0), \ldots, C_k(0), \ldots, C_K(0)\}, \\ \cup_{k=1}^{K} C_k(0) &= \{1, 2, \ldots, M\}, \\ C_i(0) \cap C_j(0) &= \emptyset \ for \ i \neq j. \end{aligned} \tag{10}$$

Correspondingly, the clustering membership for M customers is determined as

$$\begin{aligned} \boldsymbol{N}(0) &= \{\mathrm{n}_1(0), \mathrm{n}_2(0), \ldots, n_m(0), \ldots, \mathrm{n}_M(0)\} \\ n_m(0) &= k | m \in C_k, k \in [1, 2, \ldots, K] \end{aligned} \tag{11}$$

The parameters $\boldsymbol{\varphi}^{C_k(0)}$ for forecasting model $F^{C_k(0)}$ are established over the resultant clusters in (10) by fitting the aggregated load profile $\boldsymbol{S}^{C_k(0)}$ of cluster $C_k(0)$ as in (13).

$$\boldsymbol{S}^{C_k(0)} = \sum_{m \in C_k(0)} \boldsymbol{y}_m, \ C_k(0) \in \boldsymbol{C}(0) \tag{12}$$

$$\begin{aligned} \boldsymbol{\varphi}^{C_k(0)} = arg\min \left| F^{C_k(0)}\left(\widehat{\boldsymbol{S}^{C_k(0)}} \middle| \boldsymbol{S}^{C_k(0)}; \boldsymbol{\varphi}^{C_k(0)}\right) - \boldsymbol{S}^{C_k(0)} \right|, \\ C_k(0) \in \boldsymbol{C}(0) \end{aligned} \tag{13}$$

2) Feedback Mechanism: Once we have the initial forecasting models $\boldsymbol{F}^{(0)} = \{F^{C_1(0)}, F^{C_2(0)}, \ldots, F^{C_k(0)}, \ldots, F^{C_K(0)}\}$ for $\boldsymbol{C}(0)$, each customer is tested on all forecasting models to evaluate its fitness of each cluster.

To avoid the over-fitting problem, a separate validation dataset is used to test the customer's fitness over the period $u \in \{1, 2, \ldots, U\}, \boldsymbol{U} \cap \boldsymbol{T} = \emptyset$. The model fitness for customer m on $F^{C_k(0)}$ is denoted as $\varepsilon_m^{C_k(0)}$. It is the sum MAE of forecasts on m generated by the established model $F^{C_k(0)}(\boldsymbol{\varphi}^{C_k(0)})$ over the validation period as in (15).

$$\begin{aligned} \hat{y}_{m,u} &= F^{C_k(0)}\left(\boldsymbol{y}_m; \boldsymbol{\varphi}^{C_k(0)}\right), \\ u &\in \boldsymbol{U}, \ C_k \in \boldsymbol{C}, \ m \in M \end{aligned} \tag{14}$$

$$\begin{aligned} \varepsilon_m^{C_k(0)} &= \sum_{u=1}^{U} \left| y_{m,u} - \hat{y}_{m,u} \right|, \\ C_k &\in \boldsymbol{C}, \ m \in M \end{aligned} \tag{15}$$

Where $\hat{y}_{m,u}$ is the predicted electricity consumption of m by $F^{C_k(0)}(\boldsymbol{\varphi}^{C_k(0)})$ of at time u.

The established clusters and forecasting models are further developed through the feedback mechanism. The feedback signal is designed as the customers' fitness on each forecasting model. For customer m, its fitness on K forecasting models is $\varepsilon_m^{(0)} = (\varepsilon_m^{C_1(0)}, \varepsilon_m^{C_2(0)}, \ldots, \varepsilon_m^{C_K(0)})$. For the total M customers, we have the matrix of model fitness as $\boldsymbol{E} = \{(\varepsilon_m^{C_k(0)})\} \in \mathbb{R}_+^{M \times K}$. $\boldsymbol{E}$ is the feedback signal returned to the clustering stage for the re-assignment of customers.

3) Updating Clustering Membership: The 3rd step is to update the customer membership $\boldsymbol{N}$ based on $\boldsymbol{E}$. The criterion is to allocate customers to the cluster with the best model fitness.

Assign customer m to the cluster $\mathrm{n}_m(1)$ with the minimum forecasting error $\varepsilon_{m,n_m(0)}$:

$$\mathrm{n}_m(1):\ \varepsilon_{m,n_m(0)} = \operatorname{argmin} \varepsilon_m^{(0)} \tag{16}$$

Once the re-assignment for all customers is complete, the clustering membership is updated as $\boldsymbol{N}(1) = \{\mathrm{n}_1(1), \mathrm{n}_2(1), \ldots, n_m(1), \ldots, \mathrm{n}_M(1)\}$ and a new cluster set is generated as $\boldsymbol{C}(1) = \{C_1(1), C_2(1), \ldots, C_k(1), \ldots, C_K(1)\}$.

4) Updating Forecasting Models : The 4th step is to update the forecasting models based on the new cluster set $\boldsymbol{C}(1)$, following the same procedure as (12) and (13). The new set of forecasting models is established as $\boldsymbol{F}(1) = \{F^{C_1(1)}(\boldsymbol{S}^{C_1(1)}; \boldsymbol{\varphi}^{C_1(1)}), \ldots, F^{C_k(1)}(\boldsymbol{S}^{C_K(1)}; \boldsymbol{\varphi}^{C_K(2)})\}$.

5) Stopping Criterion: Step 2) - 4) would be repeated until the following criterion is met:

$$\begin{aligned} &\lambda(\mathrm{r}) < \zeta, \ \text{where } \lambda(\mathrm{r}) = |\boldsymbol{N}(r) - \boldsymbol{N}(r-1)| \\ &\text{Or } \mathrm{r} = r_{max} \end{aligned} \tag{17}$$

Where r is the number of iterations; $\lambda(r)$ is the difference of clustering memberships between the r-th and the $(r-1)$-th iteration; $\mathbf{N}(r)$ is the clustering memberships for the r-th iteration; ζ is the threshold set up at the beginning of the experiment; r_{max} is the maximum iterations. Once the criterion is met, it means that there are less than ζ switches among clusters or the computing capacity is reached. Theoretically the value of threshold is $\zeta = 1$ and the maximum number of iterations r_{max} is infinity. In this way, there would be no switches among clusters when the process terminated and convergence is guaranteed. In practice, ζ and r_{max} can be adjusted by the user according to their data size and computation limits.

B. 'Trim and Merge' of Clusters

The other advantage of the proposed method is to automatically adjust the number of clusters K by a 'Trim and Merge' strategy, without any requirements for pre-knowledge or experience.

Assuming $K = 2$ is set at the initialization stage, the two clusters are $C_\tau(0)$ and $C_\upsilon(0)$ with $|C_\tau(r)| + |C_\upsilon(r)| = M$, where $|C_k|$ represents the cardinality of cluster C_k. Over the iterative process, $|C_\tau(r)|$ is grown larger while $|C_\upsilon(r)|$ is the opposite. At the R-th iteration, it is observed that $C_\tau(R)$ is left with full members $|C_\tau(R)| = M$ and the other cluster $C_\upsilon(R)$ is empty $|C_\upsilon(R)| = 0$. It could be because that the forecasting model $F^{C_\tau(r)}(\boldsymbol{\varphi}^{C_\tau(r)})$ of cluster $C_\tau(r)$ is trained more and more general with more samples being assigned to it; $F^{C_\upsilon(r)}(\boldsymbol{\varphi}^{C_\upsilon(r)})$, on the contrary, tends to be over-fitted with limited samples on the training dataset and hence, $F^{C_\upsilon(r)}(\boldsymbol{\varphi}^{C_\upsilon(r)})$ could not deliver promising performance on the validation dataset. As a result, $\varepsilon_m^{C_\tau(r)} < \varepsilon_m^{C_\upsilon(r)}$ for $m \in C_\upsilon(r)$, it means that customers originally in the C_υ would be assigned to C_τ for the next iteration. Hence, $C_\upsilon(R)$ is trimmed and merged with $C_\tau(R)$ under the situation $|C_\upsilon(R)| = 0$. The forecasting model $F^{C_\upsilon(R)}(\boldsymbol{\varphi}^{C_\upsilon(R)})$ of $C_\upsilon(R)$ would not be updated for the following iteration but remained to be selected for the next assignment.

Under the 'trim and merge' strategy, the initial value of K needs to be set as a relatively large value at the initialization stage, such as $K = M$. Throughout the iterative process, the number of forecasting models would remain as K but the number of clusters would be trimmed and merged to its optimum value.

Comparing to the clustering-based and ensemble-based methods, the proposed method is a closed-loop system aligning the overall objective – minimizing the forecasting errors. The objective is executed throughout the iterative process. The proposed CLC method also helps to automatically identify the optimal number of clusters K and determine the clustering membership $\boldsymbol{N}$.

III. Experiments

The proposed method is validated on both simulated and real datasets to perform day-ahead forecasting, as one of the typical short-term forecasting scenarios. The application on the simulated dataset is considered as the case with the number of clusters K given while the application on the real datasets is considered as the case without any pre-knowledge or experience on K.

A. Simulated Data

1) Data Description: The simulated time-series is constructed by four components, including seasonality, trend, noise and a temperature-related component [8].

Seasonality Component: Load profiles normally represent daily, weekly and yearly cycles. To simplify the problem, only daily cycles c_t are considered in this section and implemented through a $\sin(t)$ function.

$$c_t = |\sin(\pi/24 * t)| \tag{18}$$

Trend Component: The trend component d_t is designed to simulate the moderate increasing/decreasing electricity consumption over a period. θ is the exponential index; φ_1 and φ_2 are the coefficients of t and t^θ; φ_0 is a constant.

$$d_t = \varphi_2 * t^\theta + \varphi_1 * t + \varphi_0 \tag{19}$$

Noise: The white noise ω_t is designed to simulate the volatility and uncertainty in household load profiles, which can be greatly impacted by various factors like holidays, events, *etc*. It is simulated by the ARIMA (0, 0, 0) model.

Besides the three components above, as electricity consumption such as heating and cooling devices is greatly impacted by temperature, a temperature-related component T_t is added up to the simulated time series.

The simulated time series is constructed by a linear combination of the above four components, with their individual weight to simulate different levels of impact contributed to electricity consumption.

$$\mathrm{y_t} = \beta_1 * \mathrm{c_t} + \beta_2 * \mathrm{d_t} + \beta_3 * \omega_\mathrm{t} + \beta_4 * \mathrm{T_t} \tag{20}$$

where y_t is the electricity consumption at time t and β_i is the weight coefficients of components.

2) Implementation: The time interval of the simulated data is consistent with smart meter data in real life as half-hour. The total length of the dataset is 100 days, including 4800 sample points.

Three latent classes ($K = 3$) of size 50 each are constructed:

$$\text{Class1 } y_t = \left(7 * 10^{-3} * t + 8\right) + d_t + \omega_t + T_t \tag{21}$$

$$\text{Class2 } y_t = \left(0.35 * t^{0.5} + 8\right) + d_t + \omega_t + T_t \tag{22}$$

$$\text{Class3 } y_t = \left(7 * 10^{-7} * t^2 - 2 * 10^{-4} * \mathrm{t} + 20\right) + d_t + \omega_t + T_t \tag{23}$$

With $\beta_1 = \beta_2 = \beta_4$, the three classes are designed as only different from the trend component. The coefficients of white noise are randomly generated $\beta_3 \in (9,\ 10)$ to construct more volatile datasets.

To implement day-ahead forecasting, the inputs include the electricity consumption and temperature value at the same time interval the day before the forecasting point. For example, inputs are y_{t-48} and T_{t-48} to predict y_t; the inputs also include a time indicator $t = [1, 2, \ldots, 4800]$ with its exponential forms as t^2 and $t^{0.5}$. The initial number of clusters K is set as a relatively large value as 10 to evaluate the 'Trim and merge' capability of the method. The maximum iteration and the threshold are set as $r_{max} = 100$ and $\zeta = 1$, respectively. The overall experiment is repeated for 10 times to ensure that the results are independent of ω_t.

B. Real Data

1) Data Description: Four real datasets are involved in this study, including smart metering, photovoltaic (PV) generation, temperature and solar radiation. All datasets are taking the form of half-hourly measurements from Jan. 06, 2012 to Nov. 13, 2012, including 313 days.

Moreover, datasets are then transformed into hourly intervals to assess the generalization capability of the proposed method to different cases.

- *Smart Metering:* The study utilizes anonymized electricity consumption dataset from the Irish Commission for Energy Regulation (CER) in the Smart Metering Electricity Customer Behavior Trials (CBTs) [16]. The trial covers over 5,000 Irish residential customers and businesses to implement a cost-benefit analysis for a national rollout. Among different types of customers, the 1-E-E type is used as they are the representative residential households with a controlled stimulus (E) and controlled tariff (E) and flat-billed rate.
- *PV Generation:* The PV dataset is within the South Wales area from LV Network Templates, jointly commissioned by Western Power Distribution in the U.K. and the U.K.'s regulator – the Office of Gas and Electricity Markets.
- *Temperature:* The temperature dataset is accessed from the Irish Meteorological Service [17].
- *Solar Radiation:* The solar radiation dataset is from Cardiff station under the Photovoltaic Geographical Information System of European Commission [18].

2) Implementation: With the increasing penetration of renewable energy, more customers are shifting from consumers to prosumers, who can produce and consume energy themselves. To simulate this situation, the case is designed as a group of 400 customers, including both traditional consumers and prosumers. The proportions of traditional consumers and prosumers are $(0.905, 0.095)$, respectively. The load profile of prosumers is generated through integrating 1-E-E load profiles from CBTs with PV generation profiles from LV Network Templates. PV generation is deducted from the customer's electricity consumption. The values of electricity consumption less than zero are taken as zero after the subtraction.

Three categories of inputs are considered for the forecasting model:

- *Calendar Variables:* the hour of the day, the day of the week;
- *Weather Information:* temperature and irradiation value at the same time interval the day before the forecasting point;
- *Historical Electricity Consumption:* electricity consumption at the same time interval the day before the forecasting point with its neighbouring points. Taking dataset at 30-minute intervals as an example, inputs are $y_{t-47}, y_{t-48}, y_{t-49}$ to predict y_{t+1}.

C. Comparison Study

To illustrate the effectiveness of the proposed method, it is compared with classical clustering-based forecasting methods including K-means and Gaussian mixture model (GMM); ensemble-based forecasting methods, bottom-up forecasting method and top-down forecasting method.

1) K-Means Based Methods: K-means is one of the representative partition-based clustering methods. It attempts to separate samples by minimizing the variance within clusters and has seen wide applications in clustering load profiles [15], [19].

2) Gaussian Mixture Model (GMM) Based Methods: GMM is one of the density-based clustering methods. It assumes samples within a cluster are from the same Gaussian distributions and implements the expectation-maximization algorithm to fit the mixture models [20].

3) Ensemble-Based Methods: The ensemble technique compared in this study is following [19]. Groups of forecasts are generated by varying the number of clusters and the final results are taken as the weighted average of all groups of forecasts.

4) Bottom-Up Forecasting Method: Under the bottom-up forecasting method, each smart metering dataset is fitted independently on an individual forecasting model. The final results are obtained by aggregating all individual forecasts.

5) Top-Down Forecasting Method: Under the top-down forecasting method, the results are obtained through directly conducting forecasts on the aggregated load profile.

All methods are trained on the same dataset and tested over the same period and time horizon. The selection of the forecasting model is out of scope for this paper. For simplicity, multiple linear regression (MLR) is selected as one of the

TABLE I
COMPARISON OF FORECASTING PERFORMANCE ON THE SIMULATED DATASET

Method	MAE	MAPE (%)	Improvement by CLC in terms of MAPE
a) K-means	200.66	4.33	22.86%
b) GMM	187.03	3.79	11.87%
c) Ensemble	193.26	3.96	15.66%
d) Bottom-up	207.46	4.41	24.26%
e) Top-down	212.34	4.57	26.91%
CLC	172.47	3.34	/

TABLE II
DIEBOLD-MARIANO TEST ON THE SIMULATED DATASET

$\lvert DM\rvert$ / Method	MAE	MAPE
K-MEANS	2.63	3.34
GMM	2.78	3.64
Ensemble	2.34	3.02
Bottom-up	2.09	2.35
Top-down	5.18	6.36

classical models in regression analysis. It attempts to express the response of the forecasting variable by a linear combination of a set of predictors and a constant term. The impact of each predictor on the forecasting variable is expressed by its individual coefficient, which can be derived through the least-squares method. MLR has been utilized in many studies as a benchmark model [21]–[23]. Meanwhile, it can be applied to all clustering methods and is fast for large amounts of experiments. MLR is implemented by lm() function in RStudio [24].

Two representative forecasting metrics are employed to measure the forecasting performance of different methods as mean absolute error (MAE) and mean absolute percentage error (MAPE). Due to the complexities in dealing with sampling uncertainties and correlations present in forecast errors, Diebold-Mariano (DM) test is further employed to determine whether forecasts are significantly different [25]. The null hypothesis is that the two methods have the same forecast accuracy. The equality of forecast accuracy can be investigated by computing the *DM* statistic. The null hypothesis of no difference would be rejected if the DM statistic falls within $|DM| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the z-value from the standard normal table corresponding to $1/2$ of the desired α level.

D. Training, Validation and Testing Datasets Split

The whole dataset is split into three subsets: training, validation and testing. The training subset is for the initial clustering and the establishment of forecasting models. The validation subset is for the re-assignment of the customers, to prevent the issue of over-fitting. The testing subset is for the method evaluation and comparison study among different methods. The proportion of each subset is set as 72%, 8% and 10% respectively for all methods.

IV. RESULTS AND DISCUSSIONS

In this section, we present the experimental results on both simulated datasets and real datasets. The results are compared with the classical benchmark models regarding their clustering and forecasting performance.

A. Performance on Simulated Datasets

1) Forecasting Performance: TABLE I is the comparison of forecasting performance in terms of mean absolute percentage error (MAPE) and MAE on the simulated dataset, which are taken as their average value over ten repeated experiments. The method index is consistent with Section III. The performance of K-means and GMM based method is taken with $K = 3$ given. It can be concluded that the proposed CLC method delivers the best performance on the simulated dataset with the smallest MAE and MAPE. It outperforms the top-down method the most by 26.91% in terms of MAPE, which could be due to the information loss when aggregating the time series. The CLC improves the bottom-up method the 2nd most by 24.26%, which can be due to the volatility and unpredictability of the individual series. The proposed CLC outperforms K-means and GMM based method by 22.86% and 11.87%, respectively. The reason could be that K-means and GMM can not effectively partition samples in terms of forecasting accuracy. And the performance of the ensemble-based method is improved by CLC by 15.66%. Though it is expected to offset the effect from identifying the optimal K through ensemble techniques, it does not solve the problem of the mismatched problem.

TABLE II is the results of Diebold-Mariano test on the simulated dataset. The test is taken to investigate whether the forecasting performance of the proposed CLC method is significantly different from the benchmark methods. Taking the significance level of the test as $\alpha = 0.05$, the condition to reject the null hypothesis of no difference is $|DM| > 1.96$. TABLE II shows that all of the tests have rejected the null hypothesis and hence conforms to the results in TABLE I.

The estimated parameters of the forecasting models resulted from different clustering methods are shown in TABLE III. The first row is the inputs. To implement day-ahead forecasting, the inputs for electricity consumption and temperature use the same time interval the day before forecasting as y_{t-48} and T_{t-48}, respectively. The three designed classes are different from the trend component as in (21)-(23). The key to assessing the capability in estimating parameters is the weight scale for $t, t^{0.5}, t^2$ in three classes. Class 1 takes the original form of t; class 2 takes the form of $t^{0.5}$ and class 3 includes both t and t^2. As for the proposed CLC method, the estimated parameters for y_{t-48} and T_{t-48} are almost the same for three classes; the scale of the parameters for $t, t^{0.5}, t^2$ are consistent with the designed classes; the estimated intercepts of the three models are close to the designed values. As for the estimated parameters of K-means and GMM, the weights for y_{t-48} and T_{t-48} are different among classes; the scale of coefficients for $t, t^{0.5}, t^2$ could not be the key to differentiate three classes. Hence, the capability of the proposed CLC method in estimating parameters of forecasting models could be one of the reasons enhancing the overall forecasting performance.

2) Clustering Performance: The initial number of clusters for CLC is set as 10 shrunk to 3 after 12 iterations. It is

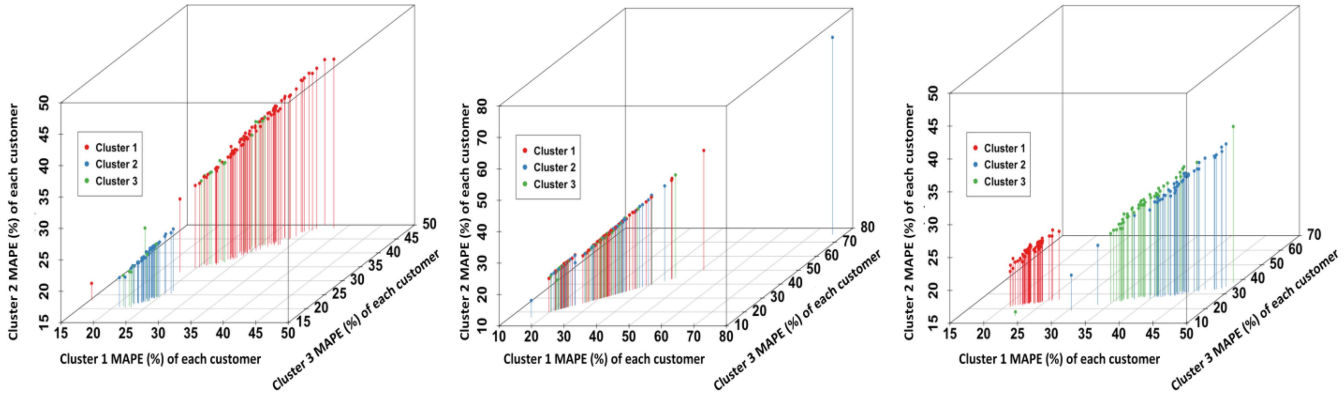


Fig. 6. MAPE (%) of all series on different forecasting models clustered by a) K-means; b) GMM; c) CLC.

TABLE III
ESTIMATED PARAMETERS OF THE MLR MODELS RESULTED FROM DIFFERENT CLUSTERING METHODS IN THE APPLICATION OF SIMULATED DATASET

Method	Input	y_{t-48}	t	T_{t-48}	t^2	$t^{0.5}$
CLC	Class1	$2.00 * 10^{-2}$	$7.29 * 10^{-3}$	$2.21 * 10^{0}$	$5.13* 10^{-9}$	$-2.28 * 10^{-2}$
	Class2	$2.18 * 10^{-2}$	$4.36 * 10^{-4}$	$2.21 * 10^{0}$	$5.53* 10^{-9}$	$3.19 * 10^{-1}$
	Class3	$2.22 * 10^{-2}$	$2.00 * 10^{-4}$	$2.21 * 10^{0}$	$6.94* 10^{-7}$	$-2.13 * 10^{-2}$
K-means	Class1	$1.72 * 10^{-2}$	$7.28 * 10^{-3}$	$2.19 * 10^{0}$	$9.73 * 10^{-9}$	$-2.20 * 10^{-2}$
	Class2	$1.91 * 10^{-2}$	$2.82 * 10^{-4}$	$2.19 * 10^{0}$	$3.56 * 10^{-7}$	$1.50 * 10^{-1}$
	Class3	$-1.51 * 10^{-3}$	$4.39 * 10^{-3}$	$3.47 * 10^{0}$	$-6.59 * 10^{-9}$	$4.71 * 10^{-2}$
GMM	Class1	$2.43 * 10^{-2}$	$2.50 * 10^{-3}$	$1.66 * 10^{0}$	$2.27 * 10^{-7}$	$9.33 * 10^{-2}$
	Class2	$-7.20 * 10^{-3}$	$-5.89 * 10^{-4}$	$2.71 * 10^{0}$	$4.76 * 10^{-7}$	$1.93 * 10^{-1}$
	Class3	$-1.51 * 10^{-3}$	$4.39 * 10^{-3}$	$3.47 * 10^{0}$	$-6.59 * 10^{-9}$	$4.71 * 10^{-2}$

consistent with the designed number of classes. It capitalizes its ability of 'trim and merge' to search for the optimal number of clusters automatically.

Fig. 6 describes the clustering results in terms of sample fitness on models. Each sub-figure is for one clustering method under the condition $K = 3$ as a) K-means; 2) GMM; 3) CLC, respectively. The three axes are for three clusters. Each axis represents the distribution of samples' MAPE tested on its corresponding forecasting model. There are 150 points in total in each sub-figure and each point represents one sample. Each sample is tested on three forecasting models and its three-dimensional coordinates are their corresponding MAPE values. Each colour represents one cluster.

As for K-means in Fig. 6(a), one large group in red and one small group in blue can be first identified. Samples in green as Cluster 3 are scattered over the distribution without any distinguish patterns. Compared with K-means, the clustering results of GMM in Fig. 6(b) represent a more disorganized pattern. Samples from different clusters are mixed up together. For

TABLE IV
CLUSTER SIZES ON THE SIMULATED DATASET

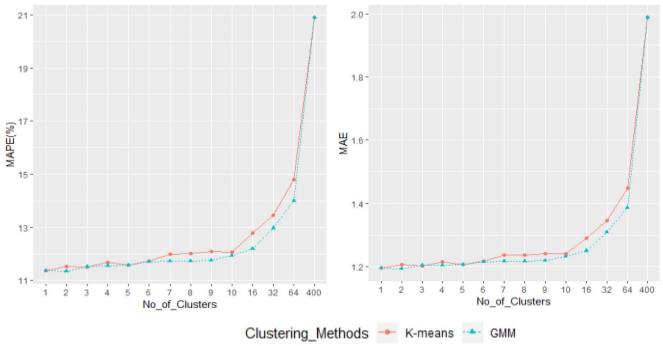
Method \ Cluster	1	2	3	Clustering Accuracy
K-means	86	43	21	76.00%
GMM	66	45	39	89.33%
CLC	50	50	50	100.00%

both K-means and GMM, there are circumstances that samples are not assigned to their best-fitted models, i.e., smallest MAPE. Fig. 6(c) is the clustering results of the proposed CLC method and three distinct groups can be identified. Most samples are assigned to the cluster with the best model fitness. The ability of the proposed method in clustering samples in terms of their model fitness does help to advance the overall forecasting accuracy.

TABLE IV represents the cluster sizes and clustering accuracy of three clustering methods. The index of the cluster as $[1, 2, 3]$ is based on the descending order of the cluster size. The clustering accuracy is calculated as the percentage of the number of correctly-clustered samples in total samples. The results show that the size for three clusters are 50 and all samples are clustered correctly by the proposed CLC method. The clustering accuracy for K-means and GMM is 76.00% and 89.33%, respectively. Overall, the 'trim and merge' strategy helps CLC to search for the optimal number of clusters automatically and the optimization with an aligned objective improves the overall forecasting performance.

B. Performance on Real Datasets

1) Forecasting Performance: For clustering-based methods like a) K-means and b) GMM, the optimal number of clusters K needs to be first identified. It is performed on the validation dataset to provide the possible values on optimal K. Fig. 7 shows the effects of the number of clusters K on the forecasting performance for the case of 30-minute interval. Sub-figure (a) and (b) are under two different evaluation metrics as MAPE and MAE. The X-axis is the number of clusters taken as $\boldsymbol{K} = [1{:}10,\ 16,\ 32,\ 64,\ 400]$ while the Y-axis is the value of the evaluation metrics taken as the average value over the validation period. The red line is for K-means



(a) MAPE (b) MAE

Fig. 7. Effects of the number of clusters on MAPE and MAE with K-means and GMM on real dataset at 30-minute intervals.

TABLE V
FORECASTING PERFORMANCE COMPARISON ON THE REAL DATASET AT 30-MIN INTERVALS

Method Index	MAE	MAPE (%)	Improvement by CLC in terms of MAPE
a)K-means	1.30	12.46	19.90
b)GMM	1.28	12.23	18.40
c)Ensemble	1.38	13.65	26.89
d)Bottom-up	1.99	20.88	52.20
e)Top-down	1.30	12.46	19.90
CLC	0.99	9.98	/

while the blue line is for GMM. The overall trend for both methods on MAPE and MAE is increasing with the number of clusters. The optimal number of clusters for K-means is $K = 1$, which makes it the same as the top-down method. The optimal number of clusters for GMM is $K = 2$. When $K = 400$, it is the bottom-up method. Both MAPE and MAE reach their maximum value, which could be due to the volatility and unpredictability of the load profile at smart-metering levels. Hence, the optimal K on the test dataset is set as $K= 1$ for K-means and $K= 2$ for GMM. Similar procedure is also conducted to the dataset at hourly intervals. The optimal K for K-means and GMM is $K = 10$ and $K = 5$, respectively.

TABLE V is the comparison of forecasting performance among different forecasting methods in terms of MAE and MAPE. The index number of each method is consistent with the previous section. The MAE and MAPE for a) K-means and b) GMM based forecasting methods are taken as their best performance case under their optimal number of clusters as $K = 1$ and $K = 2$, respectively, which makes K-means the same as top-down method. It can be seen that the proposed method delivers the best performance in this case by the minimum MAPE and MAE. It improves K-means/Top-down by 19.90% and GMM by 18.40% in terms of MAPE. It could be due to the existence of outliers or the curse of dimensionality, which drags the clustering performance of K-means and GMM. CLC improves the bottom-up method by 52.20 % due to the volatility and unpredictability of the load profile at the meter level. The ensemble-based method, as a weighted average result of varies K, is improved by CLC by 26.89%.

Similar results can be obtained for the experiment on the real dataset at 60-minute intervals. TABLE VI is the comparison of forecasting performance. The proposed CLC method also delivers the best performance in this case by outperforming top-down, GMM and K-means-based method 12.87%, 12.60%, and 8.98%, respectively. The ensemble-based method is improved by CLC by 7.10%. CLC still improves the bottom-up method the most by 38.64%. The results prove the generalization ability of CLC on different time intervals.

TABLE VI
FORECASTING PERFORMANCE COMPARISON ON THE REAL DATASET AT 60-MIN INTERVALS

Method Index	MAE	MAPE (%)	Improvement by CLC in terms of MAPE
a)K-means	5.55	31.62	8.98
b)GMM	5.66	32.93	12.60
c)Ensemble	5.49	30.98	7.10
d)Bottom-up	6.88	46.90	38.64
e)Top-down	5.70	33.03	12.87
CLC	5.17	28.78	/

TABLE VII
DIEBOLD-MARIANO TEST ON THE REAL DATASET

$\lvert DM\rvert$ / Method	30-min		60-min	
	MAE	MAPE	MAE	MAPE
K-MEANS	9.12	9.80	11.62	10.91
GMM	2.92	4.16	18.53	13.00
Ensemble	6.34	7.25	9.05	11.12
Bottom-up	26.59	28.55	9.52	13.59
Top-down	9.12	9.80	9.96	11.97

TABLE VII is the results of DM test on the real dataset to determine whether the forecasting performances are significantly different. All of the benchmark methods are compared with the proposed CLC method. Loss differential is investigated in terms of both MAE and MAPE in terms of MAE and MAPE. Similarly, the range for DM statistic to reject the null hypothesis of no difference is $|DM| > 1.96$ taking the significance level of the test as $\alpha = 0.05$. Again, the null hypothesis has been rejected by all the tests on both datasets at 30-minute and 60-minute intervals.

2) Clustering Performance: As for the experiment on the dataset at 30-minute intervals, the initial number of clusters is set as $K = 10$. The final number of clusters is $K = 4$ after the iterative process through 'Trim and merge' strategy. The cluster sizes are $[339, 57, 3, 1]$ in the descending order. The clustering results in terms of model fitness are shown in Fig. 8. The three axes represent three clusters (A, B, C) except Cluster D which only includes one customer and each axis represents the MAE distribution of customers tested on the corresponding forecasting model. 400 points represent 400 customers. The three-dimensional coordinates of each point are the MAE values tested independently on three forecasting models. Each colour represents one cluster. The majority of samples are in orange, which has a relatively stable performance with a lower level of MAE on all three forecasting models. Samples in purple, green and blue could be more volatile as their performance on three models is less steady. Overall most samples are assigned to their best-fitted clusters in terms of their model fitness. Fig. 9 is the clustering results of GMM under the optimal $K = 2$, with one cluster containing only one sample. Failure to further partition the orange

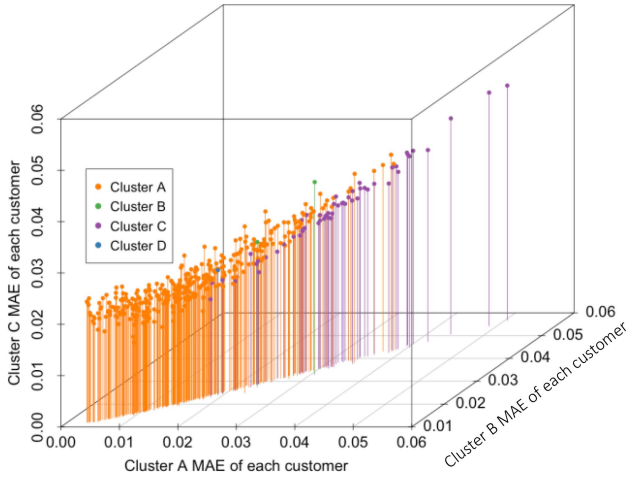


Fig. 8. CLC Clustering Results on Real Datasets in terms of Model Fitness.

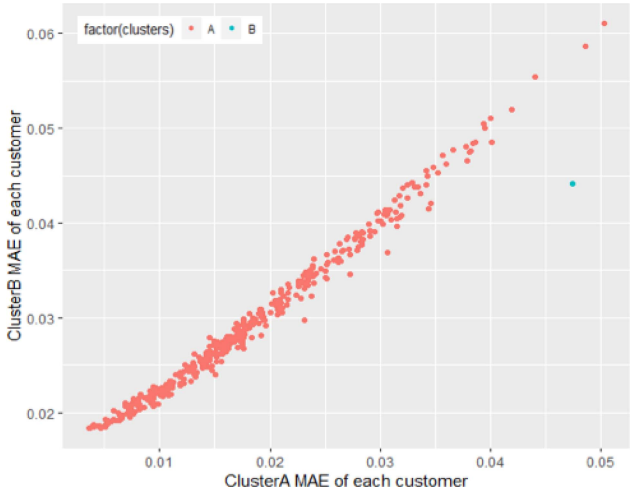


Fig. 9. GMM Clustering Results on Real Datasets in terms of Model Fitness.

groups could be the reason that GMM-based methods did not perform well on the forecasting.

Comparing to dataset at 30-minute intervals, it should be noticed the forecasting accuracy for all methods are decreased and the optimal value of K for K-means ($K = 10$), GMM ($K = 5$) and CLC is significantly increasing, when applying to the dataset at 60-minute intervals. The initial K for the proposed CLC method is set as 20 and is trimmed to 16. There could be two reasons behind it. The first is that the length of the dataset is reduced by half after transforming into 60-minute intervals, which could be not sufficient for training. The second reason could be that more activities could take place during a longer-term (i.e., 60-minute) at meter level, which results in more variations between adjacent readings and more volatilities in load profiles. Hence, more forecasting models (i.e., more clusters) are in need to fit different types of load profiles and the overall forecasting performance is depreciating.

The proposed method has a self-adaptive ability to utilize the feedback mechanism to improve the clustering and forecasting simultaneously. The cluster formation is improved by re-assigning the load patterns to the clusters which are already formed and consequently, the parameters of the corresponding forecasting model can be optimized. These capabilities can help the proposed method to partition customers making full use of the available information and hence improve the forecasting performance.

V. Conclusion

This paper for the first time proposes a closed-loop clustering method to implement hierarchical load forecasting. The closed-loop mechanism is achieved by a novel feedback mechanism linking the forecasting stage to the clustering stage. It returns the signal of model fitness and utilizes it as the clustering criterion. Hence, an iterative optimization system is established to integrate the overall objective of the process to minimize the forecasting errors. The method also contributes to automatically searching for the optimal number of clusters.

On both simulated and real datasets, we prove that the proposed method enables improvements on the forecasting performance, compared with the existing classical models. On simulated datasets, the proposed CLC method is able to identify the designed number of clusters, clustering memberships and also demonstrates its capability in estimating the parameters of the forecasting model. On real datasets, the proposed method outperforms the top-down method by 19.90%, GMM-based method by 18.40%, the ensemble-based method by 26.89%, bottom-up method by 52.20% in terms of MAPE, which validates its capability in improving the forecasting performance.

References

[1] 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, Oct. 2019.

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

[3] M. Zhou and M. Jin, "Holographic ensemble forecasting method for short-term power load," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 425–434, Jan. 2019.

[4] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.

[5] H. Jiang, Y. Zhang, E. Muljadi, J. J. Zhang, and D. W. Gao, "A short-term and high-resolution distribution system load forecasting approach using support vector regression with hybrid parameters optimization," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3341–3350, Jul. 2018.

[6] Y. Goude, R. Nedellec, and N. Kong, "Local short and middle term electricity load forecasting with semi-parametric additive models," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 440–446, Jan. 2014.

[7] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of smart meter data analytics: Applications, methodologies, and challenges," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3125–3148, May 2019.

[8] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Lexington, KY, USA: OTexts, 2013, p. 292.

[9] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I. Jennett, "Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting," *IEEE Trans. Smart Grid*, vol. 8, no. 4, pp. 1591–1598, Jul. 2017.

[10] B. Goehry, Y. Goude, P. Massart, and J. Poggi, "Aggregation of multi-scale experts for bottom-up load forecasting," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 1895–1904, May 2020.

[11] J. D. Black and W. L. W. Henson, "Hierarchical load hindcasting using reanalysis weather," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 447–455, Jan. 2014.

[12] D. Xu and Y. Tian, "A comprehensive survey of clustering algorithms," *Ann. Data Sci.*, vol. 2, no. 2, pp. 165–193, Jun. 2015.

[13] P. G. D. Silva, D. Iliæ, and S. Karnouskos, "The impact of smart grid prosumer grouping on forecasting accuracy and its benefits for local electricity market trading," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 402–410, Jan. 2014.
[14] H. Mori and A. Yuihara, "Deterministic annealing clustering for ANN-based short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 16, no. 3, pp. 545–551, Aug. 2001.
[15] F. L. Quilumba, W. 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, Mar. 2015.
[16] Irish Social Science Data Archive. (2012). *Commission for Energy Regulation (CER) Smart Metering Project*. [Online]. Available: http://www.ucd.ie/issda/data/commissionforenergyregulationcer/
[17] *MET Eireann Historical Data*. Accessed: Jun. 12, 2019. [Online]. Available: https://www.met.ie/climate/available-data/historical-data
[18] EuropeanCommission. (2012). *Photovoltaic Geographical Information System*. [Online]. Available: https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html
[19] 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.
[20] (Sep. 15, 2019). *Gaussian Mixture Models*. [Online]. Available: https://scikit-learn.org/stable/modules/mixture.html#mixture
[21] T. Hong, P. Pinson, and S. Fan, "Global energy forecasting competition 2012," *Int. J. Forecast.*, vol. 30, no. 2, pp. 357–363, Apr. 2014.
[22] T. Hong, P. Wang and H. Lee Willis, "A Naïve multiple linear regression benchmark for short term load forecasting," presented at the Power Energy Soc. Gen. Meeting, Detroit, MI, USA, Jul. 2011.
[23] T. Hong, "Short term electric load forecasting," Ph.D. dissertations, Oper. Res. Elect. Eng., North Carolina State Univ., Raleigh, NC, USA, 2010.
[24] *RStudio: Integrated Development for R*, RS Team, Boston, MA, USA, 2015. [Online]. Available: http://www.rstudio.com/
[25] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 253–263, Jul. 1995.

Chi Zhang (Graduate Student Member, IEEE) received the B.Eng. degrees in mechanical engineering from the University of Bath, Bath, U.K., and the Harbin Institute of Technology, Harbin, China, in 2014, and the M.Sc. degree in metals and energy finance from Imperial College London, London, U.K., in 2015. She is currently pursuing the Ph.D. degree with the University of Bath. Her research interests include big data and machine learning applications in power system, especially energy forecasting.

Ran Li (Member, IEEE) received the B.Eng. degrees in electrical power engineering from the University of Bath, U.K., and North China Electric Power University, Beijing, China, in 2011, and the Ph.D. degree from the University of Bath in 2014, where he is a Lecturer/Assistant Professor. His major interest is in the area of big data in power system, load profiling and forecasting, and power market and economics.