This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3142680, IEEE Access

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier ...

Smart-meter big data for load forecasting: An alternative approach to clustering

NEGIN ALEMAZKOOR[1,2,3], MAZDAK TOOTKABONI[2], ROSHANAK NATEGHI[1], and ARGHAVAN LOUHGHALAM[2]
[1]School of Industrial Engineering, Purdue University, West Lafayette, IN, USA
[2]Department of Civil and Environmental Engineering, University of Massachusetts Dartmouth, MA, USA
[3]Department of Engineering Systems and Environment, University of Virginia, VA, USA
Corresponding authors: Roshanak Nateghi & Arghavan Louhghalam.

ABSTRACT
Accurate forecasting of electricity demand is vital to the resilient management of energy systems. Recent efforts in harnessing smart-meter data to improve forecasting accuracy have primarily centered around cluster-based approaches (CBAs), where smart-meter data are grouped into a small number of clusters and separate prediction models are developed for each cluster. The cluster-based predictions are then aggregated to compute the total demand. CBAs have provided promising results compared to conventional approaches that are generally not conducive to integrating smart-meter data. However, CBAs are computationally costly and suffer from the curse of dimensionality, especially under scenarios involving smart-meter data from millions of customers. In this work, we propose an efficient reduced model approach (RMA) that leverages a novel hierarchical dimension reduction algorithm to enable the integration of fine-resolution high-dimensional smart-meter data for millions of customers in load prediction. We demonstrate the applicability of our proposed approach by using data from a utility company, based in Illinois, United States, with more than 3.7 million customers and present model performance in-terms of forecast accuracy. The proposed hierarchical dimension reduction approach enables utilizing the high-resolution data from smart-meters in a scalable manner that is not exploitable otherwise. The results shows significant improvements in forecast accuracy compared to the available approaches that either do not harness fine-resolution data or are not scalable to large-scale smart-meter big data.

INDEX TERMS Short-term load forecasting, Smart-meter data, Big data, Hierarchical dimension reduction

I. INTRODUCTION

Accurate load forecasting lies at the heart of integrated adequacy planning, and is critical for reliable and resilient operation of electric power systems [1]. Electricity supply and demand have to be matched in real time and since large-scale energy storage technology is still cost-prohibitive [2], electric utilities must carefully plan their dispatch to minimize energy loss. Moreover, accurate load prediction is an integral component of demand-side management in smart grids [3].

Electricity demand forecasting can be generally classified into short-, medium- and long-term forecasting [4]. Short-term forecasting involves predicting demand with a lead time of a few minutes to a few days, and is critical for economic dispatch [5]. Medium-term forecasts involve a prediction lead time of a few weeks to a few months, and can help electric utilities with scheduling of fuel supplies, maintenance actions and negotiation of their contracts [6]. Long-term forecasts project energy demand over a multidecadal time horizon which is crucial for capacity expansion planning under uncertain climate, policy and technological changes [7]. In this paper, we focus on short-term forecasting which allows utility companies to optimize their bidding strategies when purchasing electricity from the modern energy markets [8]. Short-term forecasting is poised to play an increasingly important role in the smart operation of the next generation power systems in the age of big-data.

The majority of studies in the literature that focus on short-term load forecasting are concerned with selecting or developing learning algorithms that result in the most accurate predictions. These include parametric and non-parametric statistical learning methods. Linear regression [9] and autoregressive integrated moving average (ARIMA) [10], are among the most widely-used parametric methods. The main advantages of these methods are their simplicity, fast training process, and high degree of interpretability [11]. With advances in statistical learning theory and computing, complex non-parametric approaches are being increasingly used to capture the non-linear patterns in demand data. Neural networks are among the most prevalent non-parametric models, with attractive features such as massive learning capability, parallelizable training process, and robustness to noise. Several studies have successfully employed neural networks for short-term load forecasting [12]–[14]. The training procedure for neural networks is, however, rather intricate. Particularly, selecting the proper architecture of (deep) neural networks and tailoring the network structure to problem complexity can be challenging. Support vector regression (SVR), another type of non-parametric algorithm, has a comparatively more straightforward training procedure and has also been extensively used for load forecasting [15]–[18]. Several other techniques such as genetic programming [19], random forests [20], Bayesian additive regression trees [21], and Gaussian process regression [22] have also been used for electricity demand prediction. Additionally, with the goal of improving forecast accuracy, researchers have proposed integration of learning methods and developed several hybrid approaches such as hybrid of ARIMA and SVM [23], and hybrid of neural networks and ARIMA [24] for short-term load forecasting.

Regardless of the choice of learning algorithm, the majority of studies on load forecasting use an aggregated model for predicting demand at spatially aggregated levels. Such aggregated models solely use previously observed *aggregated* demand as input and do not leverage fine-resolution smart-meter data [15], [16], [23], [24]. More recently, cluster-based approaches (CBAs) have been used as an alternative to aggregated models to leverage fine-resolution data [25]–[32]. CBAs involve grouping smart-meter data into clusters and developing separate prediction models for every cluster. The predicted demands across all clusters are subsequently aggregated to estimate the total demand. CBAs havev been applied to demand forecasting for both residential [25] and industrial [27] sections, and use different clustering methods including k-means [28], kernel spectral [29], and k-shape [30]. Aside from the choice of clustering algorithms, they have shown higher prediction accuracy when compared to aggregated models that are not conducive to exploiting high-resolution smart-meter data. However, a review of studies that use different CBAs for load forecasting highlights an existing gap in their applicability with large-scale data (i.e., on the order of millions customers). In fact, a close look at the most recent, well-cited studies that employ CBAs—summarized in Table 1—reveals application to limited size datasets, from a few hundreds to a few thousands customers, with more emphasis on improving the accuracy through optimal clustering [26] and using more advanced learning algorithms such as neural networks [27], deep learning [30], and hybrid methods [33]. The fact that CBAs have not been used for large datasets is mostly because of the high computational cost of widely-used clustering algorithms, such as K-means, which become computationally intractable when clustering fine-resolution smart-meter data from millions of customers is of interest. While more advanced clustering algorithms that use parallelization may alleviate the challenge to some extent, their implementation often involves complexities that prohibit their use by utility companies serving several million customers [34], [35]. Additionally, selecting the optimal number of clusters has been shown to significantly impact the accuracy of predictions [25]–[30]. The selection process, however, requires running the clustering algorithm several times for various number of clusters, a process which makes CBAs even more computationally costly.

TABLE 1: Size of datasets used for validation of cluster-based approaches in the literature.

Study	Num. of units in the dataset
Quilumba et al. (2014)	5000
Wijaya et al. (2015)	3639
Shahzadeh et al. (2015)	6000
Wang et al. (2016)	3000
Fahman et al. (2017)	3176
Fu et al. (2018)	653
Bian et al. (2020)	200

The extensive computational cost of CBAs warrants the need for an alternative approach that is conducive to big smart-meter data. In this work, we propose an efficient approach that can be readily implemented by utility companies for accurate short-term load forecasting due to (a) its scalability, rendering it an ideal approach for exploiting fine-resolution smart-meter big data, and (b) its ease of integration with various parametric and non-parametric learning algorithms.

The idea is based on hierarchically reducing data dimensionality so as to avoid overwhelming the computational and storage capacities. More specifically, we propose hierarchical principal component analysis (HPCA) to reduce the dimensionality of the input space, in a hierarchical manner, such that the reduced fine-resolution smart-meter data can be directly used as inputs to the prediction model. The development here is based on the premise that the application of classical PCA will be challenged by the sheer scale of fine-resolution smart-meter data from millions of customers, as even the declaration of the full matrix associated with such high-dimensional data could exceed the standard computational platforms available to energy systems operator. It is also in contrast to the conventional aggregated model approaches in literature that simply use spatially aggregated demand at previous time-steps as model inputs [15], [16],

[23], [24]. In fact, this work, to our best knowledge, marks the *first effort* to explore direct exploitation of smart-meter data in load forecasting through a hierarchical approach that significantly reduces the dimensionality of data in a computationally affordable manner. This work also, for the first time, uses an extremely large-scale smart-meter dataset from more than 3.7 customer units for load forecasting. Such large-scale dataset demonstrates the shortcoming of the available CBAs in dealing with smart-meter big data and further highlights the applicability of HPCA for effective dimension reduction and its efficiency and scalability when it comes to exploitation of fine-resolution smart-meter data in load forecasting. Finally we demonstrate the versatility and ease of implementation of the HPCA through integration with a variety of learning algorithms for demand prediction.

The organization of the rest of this paper is as follows. Section II reviews the aggregated model and CBAs for load forecasting, and introduces the proposed reduced model approach that harnesses smart-meter big data in short-term load forecasting. This includes the algorithmic details of HPCA in Section II-C1 and a brief review of the learning techniques used here to perform short-term forecasting in Section II-C2. By way of application, three learning methods from different classes of predictive models are selected and used to demonstrate the robustness of the proposed reduced model approach to the choice of predictive models. Section III includes a description of the data and performance measures (Section III-A) as well as an illustration of the performance of the proposed approach (Section III-B). Finally, Section IV summarizes some concluding remarks.

II. METHODOLOGY

In this section, we first review the two main conventional approaches in the literature for utility-scale demand prediction, namely, aggregated model and cluster-based approaches. We then introduce our proposed reduced model approach for efficient utility-scale demand prediction using smart-meter big data.

A. THE CLUSTER-BASED APPROACH (CBA)

Load forecasting involves predicting the total utility-scale demand at the next time-step, i.e., $u(t+1)$, given the observed utility-scale demand of previous time-steps. Consider the electricity demand at all customer units served by the utility company to be available at each time-step. A cluster-based approach typically involves grouping customer units into K clusters, as illustrated in Figure 1, and developing a separate demand prediction model for each cluster [25]–[30]. Specifically, for a given cluster c_i, a predictive model is trained to establish the relationship between the next time-step demand, $u_{c_i}(t+1)$, and recent demand observations for that cluster, $\{u_{c_i}(t), u_{c_i}(t-1), \cdots, u_{c_i}(t-r)\}$, where r indicates the number of previous observations included as model input. The total utility-level demand in the next time-step is then estimated by aggregating the demand in all clusters:

$$\tilde{u}(t+1) = \sum_{i=1}^{K} \tilde{u}_{c_i}(t+1), \tag{1}$$

where $\tilde{u}_{c_i}(t+1)$ is the predicted demand for the cluster c_i, and $\tilde{u}(t+1)$ is the predicted next time-step utility-scale demand. The number of clusters, K, can theoretically be any integer between one and the number of units served by the utility company, n. However, selecting a large K means that thousands or millions of prediction models need to be developed, a process that is computationally costly and impractical. The optimal number of clusters is often selected to be smaller than ten ($K \leq 10$) [25]–[30]. In addition, clustering the units into K clusters is computationally costly and becomes intractable, if not impossible, when the number of units is very large, i.e., when working with data from millions of customers. Therefore, demand forecasting studies in literature often use demonstrative examples with data limited to a few thousands units to investigate the applicability and assess the accuracy of CBAs [25]–[30].

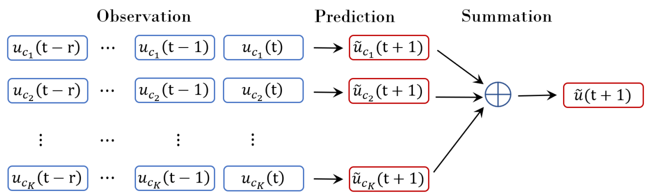


FIGURE 1: The schematic of the cluster-based prediction approach

In this work, we use large-scale smart-meter data from a utility company that serves more than 3.7 million customer units. Optimal clustering for such large volume of data is very computationally demanding if not impossible. In absence of an applicable clustering algorithm, we group customer units by their zip-codes, i.e., we consider units with similar zip-codes to be in one cluster. This probably does not provide the optimal clustering. However, we found this to be the most intuitive approach considering the large size of data and lack of clustering algorithms that can handle such large volume of data. A demand prediction model is then developed for each zip-code and the results are aggregated to predict the demand at utility-scale.

B. THE AGGREGATED MODEL APPROACH

Aggregated models consist of aggregating demand for all customer units and developing a single prediction model for the total demand [15], [16], [23], [24]. As illustrated in Figure 2, such models do not involve developing several distinct models at unit or cluster-of-units scale and are not conducive to exploiting fine-resolution smart-meter data.

C. THE REDUCED MODEL APPROACH (RMA)

Here, we propose an alternative to the aggregated model approach to allow for the efficient integration of high-resolution smart-meter data. The idea is to harness the in-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3142680, IEEE Access

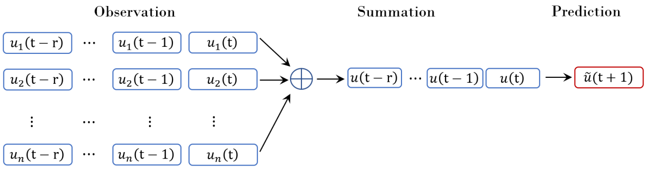


FIGURE 2: The schematic diagram of the aggregated model prediction approach

herent correlations in very high-dimensional smart-meter data and transform it to a low-dimensional space. The transformed low-dimensional data can then be directly used as inputs in a demand prediction model. Figure 3 shows the schematic of the proposed RMA. The n-dimensional data u_i, $i = 1, ..n$, is transformed to a p-dimensional space η_j, $j = 1, ..p$, with $p \ll n$. The new variables $\{\eta_1(t), \ldots \eta_1(t-r), \ldots, \eta_p(t), \ldots \eta_p(t-r)\}$ are then used as model inputs to train the predictive demand model.

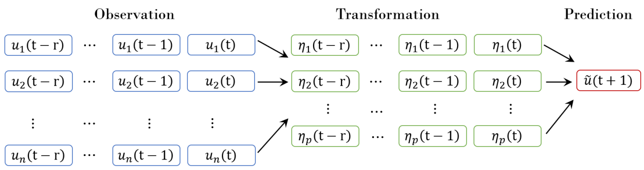


FIGURE 3: Schematic diagram of the proposed reduced model approach

1) Dimension reduction via hierarchical principal component analysis

Principal Component Analysis (PCA) is widely used for dimension reduction and mapping high-dimensional data to a low-dimensional subspace [36]. PCA looks to find an orthogonal projection of n-dimensional data onto p directions such that the largest variance comes to lie on the first direction (the first principal component), the second largest variance on the second direction, and so on. Consider $\boldsymbol{X}$ to be an $M \times N$ matrix of data, where M is the number of observations and N is the number of correlated variables. Let $\boldsymbol{w}$ be a unit vector specifying an axis in the variable space. The first principal component $\boldsymbol{w}$ is then sought as the solution of the following optimization problem:

$$\max_{\|\boldsymbol{w}\|=1} \quad \boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{w}, \tag{2}$$

The above constrained optimization problem can be turned into an unconstrained one, where the normality constraint is accounted for through a Lagrange multiplier, λ, as:

$$\max_{\boldsymbol{w},\lambda} \quad \boldsymbol{w}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{w} - \lambda(\boldsymbol{w}^T \boldsymbol{w} - 1). \tag{3}$$

Differentiating equation (3) with respect to $\boldsymbol{w}$ yields the eigenvalue problem $\boldsymbol{X}^T \boldsymbol{X} \boldsymbol{w} - \lambda \boldsymbol{w} = 0$, suggesting that λ has to be the largest eigenvalue of $\boldsymbol{X}^T \boldsymbol{X}$ and $\boldsymbol{w}$ the eigenvector associated with that. Similarly, it can be shown that the second principal component is the eigenvector associated with the second largest eigenvalue and so forth. Usually, for high-dimensional correlated data, it turns out that only the first few eigenvalues and their associated variables (computed by projecting the data onto the first few eigenvectors) is sufficient to explain most of the variance in the data.

Performing PCA has the computational complexity of $\mathcal{O}(N^3)$ [37]. In the case of high-resolution electricity demand data in this study, N is equal to the number of customer units, n, which is on the order of a few millions. Consequently, performing PCA on smart-meter big data is computationally intractable. In addition, when N is too large, even forming the full matrix $\boldsymbol{X}$, can be demanding from the point of view of memory and storage. We, therefore, propose using a Hierarchical-PCA (HPCA) that alleviates these challenges. HPCA involves (i) dividing the data into smaller subsets (ii) performing PCA on all subsets, and (iii) identifying and retaining the principal components that account for 99% of the variance in each subset. Once PCA is performed for all subsets, a data matrix that includes the selected components is formed. PCA is then performed again to further reduce the dimension of the transformed data. Algorithm 1 shows the pseudo-code for the proposed HPCA where we consider each subset to include smart-meter data for a single zip-code served by the utility company, with N_z, the total number of zip-codes in the service area. In Algorithm 1, $\boldsymbol{X}_i$ includes smart-meter data for customer units in zip-code i where $X_i(t, j)$ denotes the demand of unit j in zip-code i at time t.

Algorithm 1 Hierarchical PCA

```
Initiate η_temp to be an empty matrix
for i = 1 : N_z do                          ▷ N_z is the number of zip-codes
    Form matrix X_i                         ▷ X_i includes smart-meter data for customer units in zip-code i
    Perform PCA on X_i
    Assign η_i to include the principal components that explain 99% of variation in X_i
    η_temp = column-concatenate (η_temp, η_i)
end for
Perform PCA on η_temp
Assign η to include the principal components that explain 99% of variation in η_temp
```

As will be discussed in section III-B, HPCA allows for computationally fast and affordable dimension reduction and enables the development of a reduced predictive model, as depicted in Figure 3, that can harness high-resolution smart-meter big data. It must be noted that hierarchies are defined in HPCA rather flexibly. For data used in this work, customer units are grouped by their zip-codes. It is, therefore, intuitive to use zip-codes as a natural way for grouping data to smaller subsets. We have, however, observed that the way data is divided into subsets does not impact the accuracy of the final forecast model. The only consideration in dividing the initial data into subsets is therefore the size of subsets. For the dataset used in this work, each zip-code includes about 10^4

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3142680, IEEE Access

customer units on average, for which performing PCA takes only about 3 seconds on a personal computer. Consequently, performing the whole HPCA takes about 20 minutes, which can be further reduced to only a few minutes through parallel computing.

2) Learning methods

The RMA described in section II-C1 allows any learning technique to benefit from the high-resolution smart-meter data. In this work, we perform demand forecasting using examples from three distinct classes of learning methods to ensure the robustness of the predictions against the choice of learning technique. These include: 1) linear regression, as an example of parametric techniques, 2) SVR, as an example of deterministic non-parametric techniques, and 3) Gaussian process regression, as an example of probabilistic non-parametric techniques which, for the sake of completeness, we briefly present below. It is, however, noted that the goal is not to compare these three learning methods in terms of accuracy as performance of a particular method hinges on the characteristics of the data set. Here, the aim is to (i) evaluate the amenability of the proposed approach to various learning methods (ii) validate the independence of the observed performance on the choice of the learning method.

a: Linear Regression

Regression is simplest and yet the most widely used statistical model for prediction. In liner regression, the output y is approximated as a linear function of regressors collected in a d-dimensional vector, $\boldsymbol{x}$. That is,

$$y = b + \langle \boldsymbol{c}, \boldsymbol{x} \rangle + \epsilon, \tag{4}$$

where b is the intercept, $\boldsymbol{c}$ is the d-dimensional vector of coefficients, $\langle ., . \rangle$ denotes the dot product, and ϵ is the noise. Given a set of input and output observations, $\{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_M, y_M)\}$ the coefficients of the regression model can be easily estimated by solving the following convex minimization problem:

$$\min_{\boldsymbol{c}} \sum_{j=1}^{M} \left\| y_j - b - \langle \boldsymbol{c}, \boldsymbol{x}_j \rangle \right\|^2, \tag{5}$$

where M is the number of observations, and y_j is the output associated with the jth realization of input, $\boldsymbol{x}_j$.

b: Support Vector Regression

Consider the set of observations $\{(\boldsymbol{x}_1, y_1), \cdots, (\boldsymbol{x}_M, y_M)\}$. Support Vector Regression (SVR) with an ε margin of tolerance, ε-SVR, aims to find the flattest possible function $f(\boldsymbol{x})$ for which the deviation from all observed y_i is less than ε [38]. Consider f to be linear function, i.e, $f(\boldsymbol{x}) = b + \langle \boldsymbol{c}, \boldsymbol{x} \rangle$. The search for the flattest function with such constraint can be formulated as:

$$\begin{aligned} &\min \frac{1}{2} \|\boldsymbol{c}\|^2 \\ &\text{s.t.} \quad |y_j - b - \langle \boldsymbol{c}, \boldsymbol{x}_j \rangle| \le \varepsilon. \end{aligned} \tag{6}$$

The above optimization, however, may not be feasible as a function f that approximates all observations with error smaller than ε may not exist. To address this challenge, two slack variables are introduced [38] and (6) is rewritten as:

$$\begin{aligned} &\min \frac{1}{2} \|\boldsymbol{c}\|^2 + C \sum_{j=1}^{M} (\xi_j + \xi_j^*) \\ &\text{s.t.} \quad \begin{cases} y_j - b - \langle \boldsymbol{c}, \boldsymbol{x}_j \rangle \le \varepsilon + \xi_j \\ b + \langle \boldsymbol{c}, \boldsymbol{x}_j \rangle - y_j \le \varepsilon + \xi_j^* \\ \xi_j, \xi_j^* \ge 0, \end{cases} \end{aligned} \tag{7}$$

where constant $C > 0$ determines the trade-off between the flatness of f and the deviation of approximations from observations. In order to facilitate computationally efficient approximation of non-linear cases, SVR utilizes kernel trick. This trick replaces $\boldsymbol{x}_j$ in (7) with $\phi(\boldsymbol{x}_j)$, i.e., ϕ is a kernel function, and implicitly maps the input space into a higher dimensional space. One of the most widely used kernel functions is the Gaussian kernel which we use in this work to construct the SVR models. Readers are referred to [39] for more details.

c: Gaussian Process Regression

Gaussian Processes Regression (GPR) is a non-parametric Bayesian approach that extends multivariate Gaussian distribution to infinite dimensions. GPR has been widely used as it is analytically tractable, and has a robust and probabilistic work-flow [40]. Consider the regression problem given by:

$$y = f(\boldsymbol{x}) + \epsilon, \tag{8}$$

where $f(\boldsymbol{x})$ is the underlying regression function to be approximated, and ϵ is the noise, usually considered to be Normally distributed, i.e., $\epsilon \sim \mathcal{N}(0, \sigma_n^2)$. GP defines a probability distribution over $f(\boldsymbol{x})$, which is (often) specified by a zero mean and a covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$. One of the most commonly used covariance functions is the *squared exponential*,

$$k(\boldsymbol{x}, \boldsymbol{x}') = \sigma_f^2 \exp\left\{ \sum_{i=1}^{d} \left(\frac{-(x^{(i)} - x'^{(i)})^2}{2 l_i^2} \right) \right\} + \sigma_n^2 \delta(\boldsymbol{x}, \boldsymbol{x}'), \tag{9}$$

with σ_f, $\boldsymbol{l}$ and σ_n the hyperparameters of the covariance function, σ_f the maximum allowable variance, $\boldsymbol{l}$ the vector of length scale parameters, σ_n the measurement noise, and $\delta(x, x')$ the Kronecker delta function. Let $\boldsymbol{\theta}$ be the vector of hyperparameters for the covariance function:

$$\boldsymbol{\theta} := [\sigma_f, \boldsymbol{l}, \sigma_n]. \tag{10}$$

Given the training data, i.e., an observed input matrix, $\boldsymbol{X} := (\boldsymbol{x}_1; \ldots; \boldsymbol{x}_M)$, and observed output vector, $\boldsymbol{y}$, $\boldsymbol{\theta}$ can be estimated through maximum likelihood. Once the hyperparameters are estimated given the observed data, for a new input observation, $\boldsymbol{x}^*$, the output y^* follows

$$y^* | \boldsymbol{x}^*, \boldsymbol{X}, \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{k}^* \boldsymbol{K}^{-1} \boldsymbol{y}, k^{**} - \boldsymbol{k}^* \boldsymbol{K}^{-1} \boldsymbol{k}^{*T}), \tag{11}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3142680, IEEE Access

where

$$\boldsymbol{k}^* := k(\boldsymbol{x}^*, \boldsymbol{X}) = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}^*) & \cdots & k(\boldsymbol{x}_M, \boldsymbol{x}^*) \end{bmatrix}. \quad (12)$$

$\boldsymbol{K}$ is the $n \times n$ matrix derived from evaluating the covariance function at observed inputs:

$$\boldsymbol{K} := k(\boldsymbol{X}, \boldsymbol{X}) = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k(\boldsymbol{x}_M, \boldsymbol{x}_1) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_1, \boldsymbol{x}_M) & \cdots & k(\boldsymbol{x}_M, \boldsymbol{x}_M) \end{bmatrix}, \quad (13)$$

and $k^{**} := k(\boldsymbol{x}^*, \boldsymbol{x}^*)$. The mean of the Gaussian distribution in (11) is finally used as the best estimation for y^*.

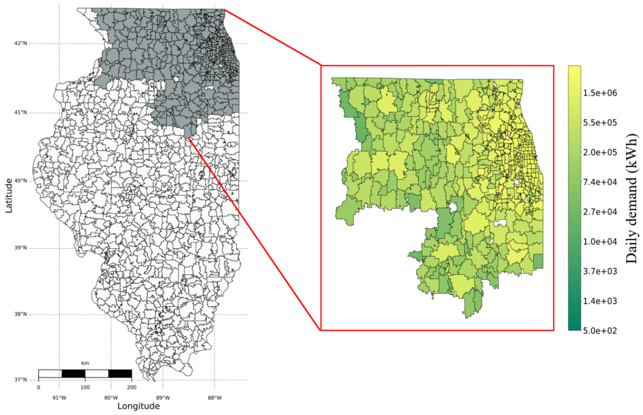


FIGURE 4: ComEd service area map. The colored area shows the ComEd territory.

III. EMPIRICAL RESULTS

In this section, we briefly introduce the data used for this study and the performance measures used for accuracy evaluation. We then discuss the empirical design used to carry out the investigation, and report the results on prediction accuracy.

TABLE 2: Mean, median, and standard deviation of 30-min demand at zip-code level in kWh across all zip-codes in ComEd's service territory.

	Mean	Median	Standard deviation
Weekdays	5705	2834	7704
Weekends	5098	2684	6571

A. DATA AND PERFORMANCE MEASURE

We use demand data procured from Commonwealth Edison, commonly known as ComEd, which is the largest electric utility in the state of Illinois. ComEd serves more than 500 zip-codes and 3.7 million customer units in Illinois; see Figure 4 for the service territory of ComEd.

The ComEd data used in this study includes 30-min demands for all units in all zip-codes in ComEd's service territory. To provide an understanding of 30-min demand in ComEd's service territory, distributions of 30-min demand at zip-code level across all zip-codes for weekdays and weekends are depicted in Figure 5. Additionally, Table 2 includes the mean, median, and standard deviation of the distributions shown in Figure 5.

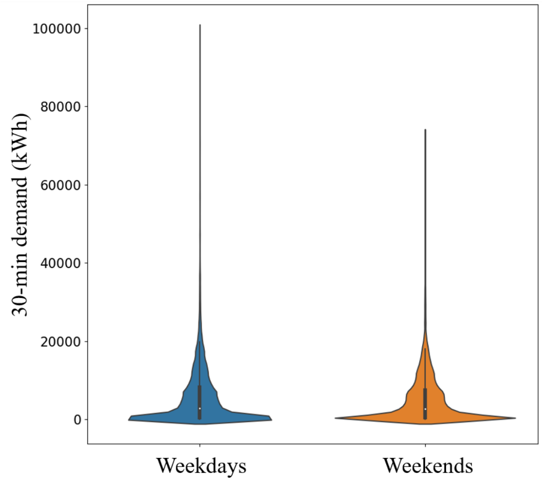
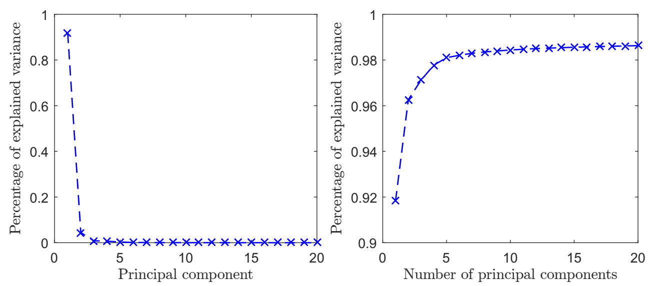


FIGURE 5: Distributions of 30-min demand at zip-code level across all zip-codes in ComEd's service territory.



(a) Explained variance by each component (b) Cumulative explained variance

FIGURE 6: Percentage of explained variance in smart-meter big data (a) by individual principal components, and (b) cumulatively.

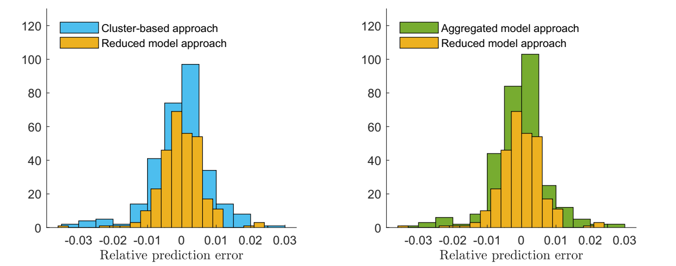
We use demand data from October 2019 as an example of an uneventful month with no major holiday or climatic extremes to ensure a fair comparison of different approaches. The data is divided into a training set and a test set. The training set includes 80% of the available data, i.e. the first 25 days of October. As shown in Figures 1–3, inputs to the models include the demand for r previous time-steps. To obtain an optimum value for r we use 80% of the data in the training set for training the predictive models with different values of r and perform validation using the remaining 20% of the data in the training set.

Finally, the remaining 20% of the data (i.e., the last 6 days of October) is used as the test set to evaluate the models' generalization performance.

In our analysis we found that adding variables that indicate time-of-day or day-of-week did not improve the prediction accuracy. Adding weather-related input variables may im-

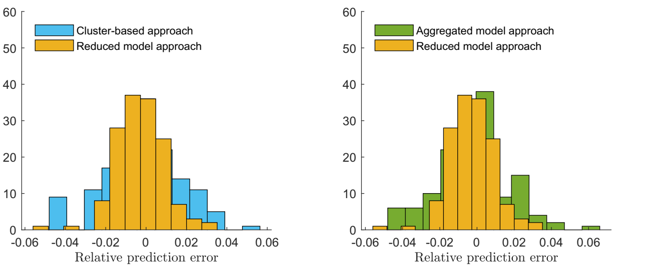
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3142680, IEEE Access

prove the accuracy but was not considered in this work since the main concern was providing an scheme that allows for effective incorporation of extremely high-dimensional smart-meter data in short-term load forecasting. We, however, note that adding weather-related inputs only slightly increases the input dimensionality and would be easily feasible if desired. Additionally, considering the substantial dimension-reduction achieved by HPCA, we found that only a few weeks of data is sufficient for training the forecast models. Reducing the required volume of training data, in fact, can be seen as another advantage of the proposed approach as it facilitates the training procedure.



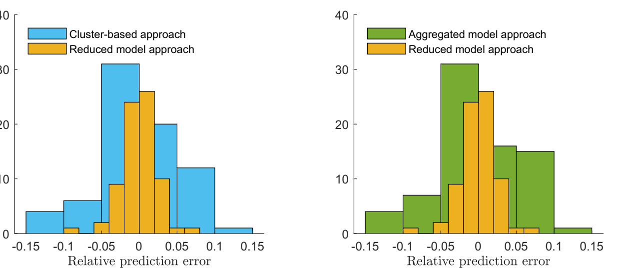
(a) Reduced model vs. cluster-based approach (b) Reduced model vs. aggregated model approach

FIGURE 7: Histogram of prediction error for 30-min prediction interval using different prediction approaches



(a) Reduced model vs. cluster-based approach (b) Reduced model vs. aggregated model approach

FIGURE 8: Histogram of prediction error for 1-hr prediction interval using different prediction approaches



(a) Reduced model vs. cluster-based approach (b) Reduced model vs. aggregated model approach

FIGURE 9: Histogram of prediction error for 2-hr prediction interval using different prediction approaches

To compare the effectiveness of different predictive modeling approaches and ensure the robustness of results against the performance measure, we use two different measures, i.e. the mean absolute percentage error and the coefficient of variation as measures of error. The mean absolute percentage error, e_{MAP}, between the vectors of observed total demand $\boldsymbol{u}$ and predicted total demand $\widetilde{\boldsymbol{u}}$ is defined as:

$$e_{\text{MAP}} = \frac{1}{N_o} \sum_{i=1}^{N_o} \frac{|u(i) - \widetilde{u}(i)|}{u(i)}, \tag{14}$$

where N_o is the number of observations in the test set. The coefficient of variation, e_{CV}, reads:

$$e_{\text{CV}} = \frac{\sqrt{\frac{1}{N_o-1} \sum_{i=1}^{N_o} (u(i) - \widetilde{u}(i))^2}}{\bar{u}}, \tag{15}$$

where $\bar{u}$ is the average total demand.

The proposed RMA outperforms both the aggregated model and CBAs due to its ability to efficiently exploit fine-resolution smart-meter data toward more accurate demand prediction. In fact, RMA consistently results in the best accuracy across different learning methods and different temporal resolutions.

B. RESULTS

As explained in Section II-C, we use HPCA to reduce the dimensionality of high-resolution smart-meter data. Figure 6a depicts the contribution of the first 20 principal components to total variance of 30-min smart-meter data. As observed in Figure 6b, the cumulative explained variance reaches a plateau with the first six principal components included, accounting for 98% of variance in the data. We thus use recent observed values, from r previous time-steps, for the first six principal components as inputs to the reduced demand prediction models. The results are then compared with aggregated model and CBAs in-terms of forecast accuracy. Table 3 summarizes coefficient of variation, e_{CV}, for the aggregated model and CBAs as well as the proposed RMA using the three learning methods. The comparison is performed for different temporal resolution, i.e. for 30-min, 1-hr, and 2-hr lead times. Since searching for the optimal number of clusters for data from 3.7 million customer units is computationally intractable, for the cluster-based approach, we group customer units data based on their zip-codes. We, however, note that in the absence of optimal clustering, mimicking the clustering process of CBAs with grouping based on some inherent properties in the data (e.g. zipcodes) results in relatively poor prediction accuracy. This is not surprising as the number of clusters has been shown to significantly impact the accuracy of CBAs [25]–[30]. In summary, CBAs by virtue of their dependence on optimal clustering, are not scalable for scenarios where smart-meter big data from millions of customers is to be dealt with.

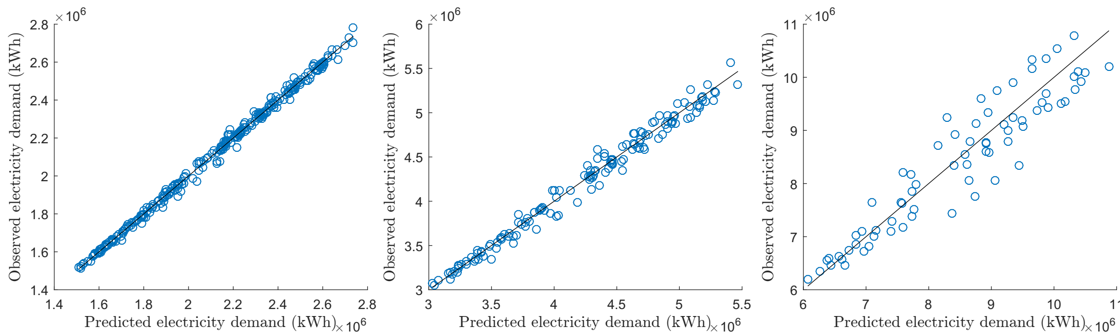
Same observation is made when the mean absolute percentage error, e_{MAP}, is used as the performance measure for different models and prediction intervals (see Table 4).

TABLE 3: Evaluated e_{CV} in % for cluster-based, aggregated model, and reduced model approaches for 30-min, 1-hr, 2-hr prediction interval using three different learning methods.

	Linear regression			Support vector regression			Gaussian process regression		
	30-min	1-hr	2-hr	30-min	1-hr	2-hr	30-min	1-hr	2-hr
Cluster-based approach	1.2	2.1	5.6	1.3	3.2	6.5	1.3	2.6	6.5
Aggregated model approach	0.8	1.9	5.5	0.8	1.9	5.8	1.0	2.3	5.6
Reduced model approach	0.6	1.2	2.5	0.8	1.6	4.4	0.6	1.4	2.4

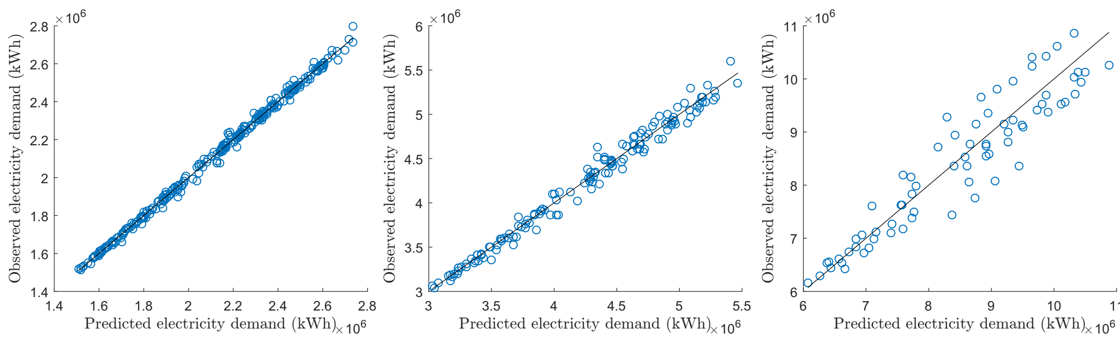
TABLE 4: Evaluated e_{MAP} in % for cluster-based, aggregated model, and reduced model approaches for 30-min, 1-hr, 2-hr prediction interval using three different prediction models.

	Linear regression			Support vector regression			Gaussian process regression		
	30-min	1-hr	2-hr	30-min	1-hr	2-hr	30-min	1-hr	2-hr
Cluster-based approach	0.7	1.5	4.4	0.9	2.1	4.9	0.8	2.2	5.4
Aggregated model approach	0.5	1.4	4.4	0.6	1.4	4.7	0.6	1.6	3.7
Reduced model approach	0.4	0.9	1.8	0.6	1.3	2.8	0.4	1.1	1.6



(a) 30-min prediction interval (b) 1-hr prediction interval (c) 2-hr prediction interval

FIGURE 10: Observed and predicted demand for different prediction intervals using the cluster-based approach



(a) 30-min prediction interval (b) 1-hr prediction interval (c) 2-hr prediction interval

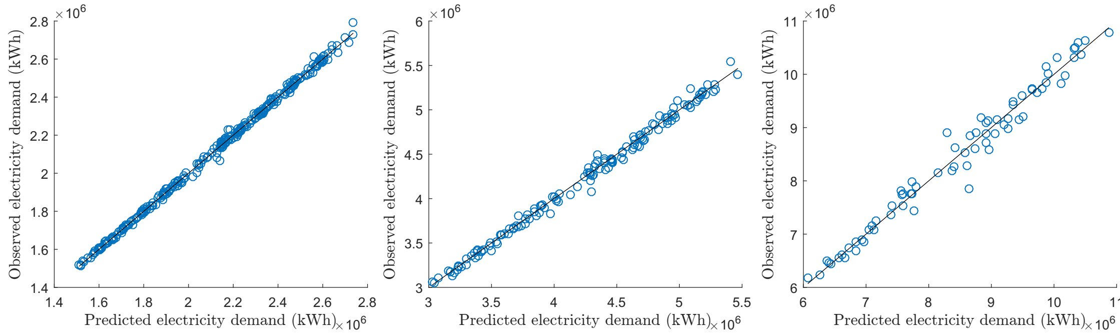
FIGURE 11: Observed and predicted demand for different prediction intervals using the aggregated model approach

This further confirms that the improved accuracy obtained by RMA is robust to the choice of both learning methods and performance measure.

We next investigate whether the accuracy improvement achieved by the proposed RMA is statistically significant. For the sake of brevity, we focus on the results obtained from the linear regression model. Figures 7, 8, and 9 compare the distributions of relative demand prediction error for different approaches for 30-min, 1-hr, and 2-hr prediction intervals, respectively. It is observed that the proposed approach results in smaller error variance compared to the aggregated model and CBAs. Assuming that error has a normal distribution, we use F-test to investigate whether the observed difference in error variance is statistically significant. It is found that, for all prediction intervals, the observed difference is statistically significant at 99% confidence level. In other words, using the proposed RMA results in statistically significant improvement in load forecast accuracy compared to aggregated model and CBAs.

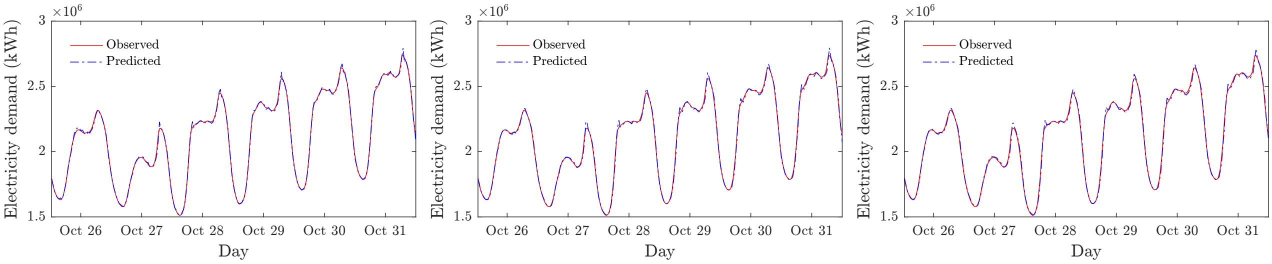
Figures 10–12 compare the observed and predicted demand for the test set (i.e., last 6 days of October) using different learning methods for 30-min, 1-hr, and 2-hr prediction intervals, respectively. It is observed that predictions tend to deviate more from the observed demand for longer prediction intervals. However, as evident in Figure 12, deviations from the observed demand are significantly smaller when the proposed RMA is used for prediction.

To better understand the significance of accuracy improve-



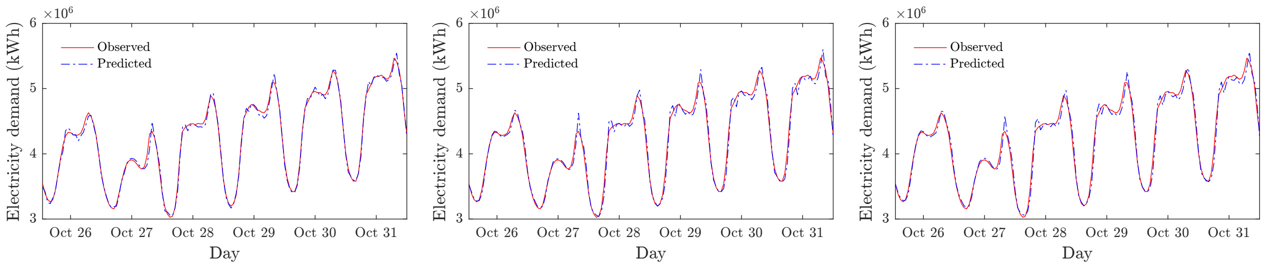
(a) 30-min prediction interval (b) 1-hr prediction interval (c) 2-hr prediction interval

FIGURE 12: Observed and predicted demand for different prediction intervals using the reduced model approach



(a) Reduced model approach (b) Aggregated model approach (c) Cluster-based approach

FIGURE 13: Observed and predicted demand for 30-min prediction interval during the last week of October 2019 using different prediction approaches



(a) Reduced model approach (b) Aggregated model approach (c) Cluster-based approach

FIGURE 14: Observed and predicted demand for 1-hr prediction interval during the last week of October 2019 using different prediction approaches

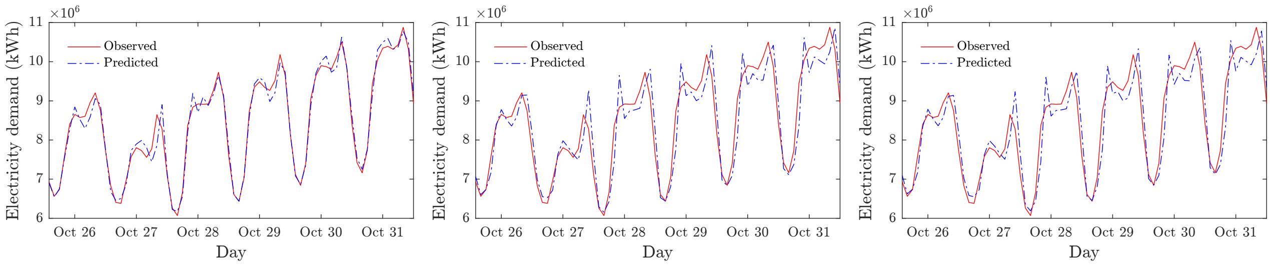
ment, Figures 13–15 compare the observed and predicted demand for the last 6 days of October 2019. We again observe that the extent of improvement in accuracy for the proposed RMA is more evident for longer prediction intervals (as we move from 13 to 15). This suggests that utility companies can significantly improve the accuracy of the load forecasts through exploiting smart-meter data using the proposed RMA, specially when the optimal operation requires longer prediction intervals.

IV. CONCLUSION

Exploiting fine-resolution smart-meter data enables more accurate short-term load forecasting. The current approaches developed to 'mine' fine-resolution data, however, are computationally expensive and not scalable to large-scale smart-meter big data. This precludes harnessing the big data revolution for more sustainable and optimal management of the grid. In this work, we propose a reduced model approach that exploits hierarchical principal component analysis (HPCA) for efficient and computationally affordable integration of smart-meter big data into short-term demand forecasting. Efficiency is achieved through transforming the high-dimensional data to a low-dimensional space hierarchically and using the transformed data as input to predict aggregate utility-scale demand.

We use large-scale smart-meter data from a utility company that serves more than 3.7 customer units to evaluate the proposed approach in terms of forecast accuracy. We find that the proposed RMA results in significantly more accurate forecasts compared to aggregated model and CBAs that are not suitable for harnessing smart-meter big data. The ability to harness high resolution data while improving the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2022.3142680, IEEE Access



(a) Reduced model approach (b) Aggregated model approach (c) Cluster-based approach

FIGURE 15: Observed and predicted demand for 2-hr prediction interval during the last week of October 2019 using different prediction approaches

forecast accuracy can help utility companies to better plan purchasing and selling electric power, load switching, and optimal system operation, thereby enhancing the reliability of the system.

In summary, we outline the key gaps in the state-of-the-art load forecasting and offer a scalable approach for more accurate and yet efficient projection of the load using large-scale smart-meter data. Our results have significant implications for achieving sustainable development goals. This is because the energy systems play a crucial role in transitioning to smart and sustainable urban systems, and access to scalable methodologies that harness the big data is essentially a prerequisite for its modernization and transition to the next generation smart grid.

ACKNOWLEDGMENT

The authors would like to acknowledge Purdue Climate Change Research Center (PCCRC), the Center for Environment (C4E) as well as the NSF grants CRISP-1832688, CMMI-1826161 and CMMI-1826155. The authors would also like to acknowledge Minsoo Choi for his help with visualizing the service area map.

REFERENCES

[1] D. Maia-Silva, R. Kumar, and R. Nateghi, "The critical role of humidity in modeling summer electricity demand across the united states," *Nature Communications*, vol. 11, no. 1, pp. 1–8, 2020.

[2] O. Schmidt, A. Hawkes, A. Gambhir, and I. Staffell, "The future cost of electrical energy storage based on experience rates," *Nature Energy*, vol. 2, no. 8, pp. 1–8, 2017.

[3] N. Li, L. Chen, and S. H. Low, "Optimal demand response based on utility maximization in power networks," in *2011 IEEE power and energy society general meeting*. IEEE, 2011, pp. 1–8.

[4] F. Ardakani and M. Ardehali, "Long-term electrical energy consumption forecasting for developing and developed economies based on different optimized models and historical data types," *Energy*, vol. 65, pp. 452–461, 2014.

[5] P. Alipour, S. Mukherjee, and R. Nateghi, "Assessing climate sensitivity of peak electricity load for resilient power systems planning and operation: A study applied to the texas region," *Energy*, vol. 185, pp. 1143–1153, 2019.

[6] F. Apadula, A. Bassini, A. Elli, and S. Scapin, "Relationships between meteorological variables and monthly electricity demand," *Applied Energy*, vol. 98, pp. 346–356, 2012.

[7] S. Mukhopadhyay and R. Nateghi, "Estimating climateâĂŤdemand nexus to support longterm adequacy planning in the energy sector," in *2017 IEEE Power & Energy Society General Meeting*. IEEE, 2017, pp. 1–5.

[8] R. Ghorani, M. Fotuhi-Firuzabad, and M. Moeini-Aghtaie, "Optimal bidding strategy of transactive agents in local energy markets," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5152–5162, 2018.

[9] G. Dudek, "Pattern-based local linear regression models for short-term load forecasting," *Electric Power Systems Research*, vol. 130, pp. 139–147, 2016.

[10] C.-M. Lee and C.-N. Ko, "Short-term load forecasting using lifting scheme and arima models," *Expert Systems with Applications*, vol. 38, no. 5, pp. 5902–5911, 2011.

[11] D. Lee and S. Derrible, "Predicting residential water demand with machine-based statistical learning," *Journal of Water Resources Planning and Management*, vol. 146, no. 1, p. 04019067, 2020.

[12] T. Peng, N. Hubele, and G. Karady, "Advancement in the application of neural networks for short-term load forecasting," *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 250–257, 1992.

[13] S. Ryu, J. Noh, and H. Kim, "Deep neural network based demand side short term load forecasting," *Energies*, vol. 10, no. 1, p. 3, 2017.

[14] A. Baliyan, K. Gaurav, and S. K. Mishra, "A review of short term load forecasting using artificial neural network models," *Procedia Computer Science*, vol. 48, pp. 121–125, 2015.

[15] E. Ceperic, V. Ceperic, and A. Baric, "A strategy for short-term load forecasting by support vector regression machines," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4356–4364, 2013.

[16] Y. Li, J. Che, and Y. Yang, "Subsampled support vector regression ensemble for short term electric load forecasting," *Energy*, vol. 164, pp. 160–170, 2018.

[17] M. Lokhandwala and R. Nateghi, "Leveraging advanced predictive analytics to assess commercial cooling load in the us," *Sustainable Production and Consumption*, vol. 14, pp. 66–81, 2018.

[18] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Applied Energy*, vol. 123, pp. 168–178, 2014.

[19] Y.-S. Lee and L.-I. Tong, "Forecasting energy consumption using a grey model improved by incorporating genetic programming," *Energy conversion and Management*, vol. 52, no. 1, pp. 147–152, 2011.

[20] A. Lahouar and J. B. H. Slama, "Random forests model for one day ahead load forecasting," in *IREC2015 The Sixth International Renewable Energy Congress*. IEEE, 2015, pp. 1–6.

[21] T. Chen, J. Lehr, O. Lavrova, and M. Martinez-Ramonz, "Distribution-level peak load prediction based on bayesian additive regression trees," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*. IEEE, 2016, pp. 1–5.

[22] M. Blum and M. Riedmiller, "Electricity demand forecasting using gaussian processes," in *Workshops at the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 2013.

[23] S. Karthika, V. Margaret, and K. Balaraman, "Hybrid short term load forecasting using arima-svm," in *2017 Innovations in Power and Advanced Computing Technologies (i-PACT)*. IEEE, 2017, pp. 1–7.

[24] J. Zhang, Y.-M. Wei, D. Li, Z. Tan, and J. Zhou, "Short term electricity load forecasting using a hybrid model," *Energy*, vol. 158, pp. 774–781, 2018.

[25] T. K. Wijaya, M. Vasirani, S. Humeau, and K. Aberer, "Cluster-based aggregate forecasting for residential electricity demand using smart meter data," in *2015 IEEE international conference on Big data (Big data)*. IEEE, 2015, pp. 879–887.

[26] X. Wang, W.-J. Lee, H. Huang, R. L. Szabados, D. Y. Wang, and P. Van Olinda, "Factors that impact the accuracy of clustering-based load

forecasting," *IEEE Transactions on Industry Applications*, vol. 52, no. 5, pp. 3625–3630, 2016.

[27] 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 Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, 2014.

[28] A. Shahzadeh, A. Khosravi, and S. Nahavandi, "Improving load forecast accuracy by clustering consumers using smart meter data," in *2015 international joint conference on neural networks (IJCNN)*. IEEE, 2015, pp. 1–7.

[29] C. Alzate and M. Sinn, "Improved electricity load forecasting via kernel spectral clustering of smart meters," in *2013 IEEE 13th International Conference on Data Mining*. IEEE, 2013, pp. 943–948.

[30] F. Fahiman, S. M. Erfani, S. Rajasegarar, M. Palaniswami, and C. Leckie, "Improving load forecasting based on deep learning and k-shape clustering," in *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2017, pp. 4134–4141.

[31] 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.

[32] H. Bian, Y. Zhong, J. Sun, and F. Shi, "Study on power consumption load forecast based on k-means clustering and fcm–bp model," *Energy Reports*, vol. 6, pp. 693–700, 2020.

[33] H. H. Aly, "A proposed intelligent short-term load forecasting hybrid models of ann, wnn and kf based on clustering techniques for smart grid," *Electric Power Systems Research*, vol. 182, p. 106191, 2020.

[34] M. S. Hidri, M. A. Zoghlami, and R. B. Ayed, "Speeding up the large-scale consensus fuzzy clustering for handling big data," *Fuzzy Sets and Systems*, vol. 348, pp. 50–74, 2018.

[35] R. Ding, Q. Wang, Y. Dang, Q. Fu, H. Zhang, and D. Zhang, "Yading: Fast clustering of large-scale time series data," *Proceedings of the VLDB Endowment*, vol. 8, no. 5, pp. 473–484, 2015.

[36] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[37] T. Elgamal and M. Hefeeda, "Analysis of pca algorithms in distributed environments," *arXiv preprint arXiv:1503.05214*, 2015.

[38] V. Vapnik, "The nature of statistical learning theory," *New York*, 1995.

[39] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.

[40] M. Raissi, H. Babaee, and G. E. Karniadakis, "Parametric gaussian process regression for big data," *Computational Mechanics*, vol. 64, no. 2, pp. 409–416, 2019.

NEGIN ALEMAZKOOR is currently an Assistant Professor in the Department of Engineering Systems and Environment at University of Virginia. In 2019, she earned her PhD in Sustainable and Resilient Infrastructure Systems program in the Department of Civil and Environmental Engineering from the University of Illinois at Urbana-Champaign. She received her Bachelor's degree in Civil Engineering from Sharif University of Technology in Tehran, Iran, and her Master's in Civil Engineering from Texas A&M University. Her current research mainly concerns developing computational tools for fast and reliable analysis of smart and resilient infrastructure systems with a focus on power and transportation systems.

MAZDAK TOOTKABONI is currently an Associate Professor in the Department of Civil and Environmental Engineering an a member of Center for Scientific Computing at University of Massachusetts Dartmouth. He earned his PhD from the Department of Civil and Systems Engineering at the Johns Hopkins University and his Bachelor's and Masters degrees in Civil Engineering from Tehran University in Tehran, Iran and Johns Hopkins University. Dr. Tootkaboni's research lies at the intersection of computational modeling, applied probability and statistics and data analytics with applications in resilient, reliable and resource efficient civil infrastructure. He is an associate member of ASCE and a member of Engineering Mechanics Institute (EMI) and its Probabilistic Methods Committee.

ROSHANAK NATEGHI is an Associate Professor of Industrial Engineering at Purdue University. Her research focuses on developing interdisciplinary methods to model the sustainability, risk and resilience of critical infrastructure under natural hazards and climate change. Prior to joining Purdue in 2015, she was an NSF Science, Engineering and Education for Sustainability Fellow, jointly appointed between Johns Hopkins University and Resources for the Future. She completed her undergraduate degree in Mechanical Engineering at Imperial College London (2006), and received her MSE (2009) and PhD (2012) degrees in Environmental Engineering at Johns Hopkins University.

ARGHAVAN LOUHGHALAM is an assistant professor in the department of Civil and Environmental Engineering with a joint appointment in Mechanical Engineering Department at University of Massachusetts, Dartmouth. Her research interests lie in the area of engineering mechanics, physics-constrained data-driven modeling, and applied statistics with particular emphasis on development of smart solutions for resilient and sustainable built environment. Prior to joining UMass Dartmouth she was a postdoctoral research associate at Massachusetts Institute of Technology. She earned her PhD from the Department of Civil and Systems Engineering at the Johns Hopkins University. She has a master's degree from University of Tehran and an undergraduate degree from Iran University of Science and Technology.