BERT-PIN: A BERT-based Framework for Recovering Missing Data Segments in Time-series Load Profiles

Yi Hu, *Graduate Student Member, IEEE*, Kai Ye, *Graduate Student Member*, *IEEE*, Hyeonjin Kim, *Graduate Student Member, IEEE*, *and* Ning Lu, *Fellow, IEEE*

***Abstract*—Restoring missing data holds paramount importance in power system analysis. Traditional recovery methods typically offer only a singular solution, lacking adaptability and depth. To bridge this gap, we introduce BERT-PIN, a pioneering approach harnessing Bidirectional Encoder Representations from Transformers for Profile Inpainting. This innovative technique enables the recovery of multiple segments of missing data by leveraging power system load and temperature profiles. Our findings demonstrate that BERT-PIN enhances accuracy by 5%-30% compared to existing techniques, showcasing its ability to restore numerous missing data segments across extended periods. We have successfully applied BERT-PIN in two critical power system applications: recovering missing data segments and estimating Conservation Voltage Reduction baselines. Serving as a versatile pre-trained model, BERT-PIN supports various downstream tasks, including classification and super-resolution, thereby reducing the necessity for extensive training data, and minimizing training time.**

Index Terms*— *Bidirectional Encoder Representations from Transformers, Conservation Voltage Reduction, Machine learning, Missing data restoration, Power System, Transformer.

NOMENCLATURE

Scalar

e	Threshold for selecting the fork points
N	The length of the time series load profile
N_{Agg}	Aggregation Level
N_m	Number of MDS in a load profile
P_{MAX}	Peak power of aggregated load profile
t_{start}	Start time of MDSs
t_{end}	End time of MDSs

Vector/Matrix

$\boldsymbol{D}$	Probability distribution matrix
$\boldsymbol{M}$	Masking vector
$\boldsymbol{X}$	A time series load profile
$\tilde{\boldsymbol{X}}$	Non-missing part in $\boldsymbol{X}$
$\boldsymbol{X}_m^{\square}$	A missing data segment in $\boldsymbol{X}$
$\hat{\boldsymbol{X}}_{m,i}^{\square}$	Estimated missing data in $\boldsymbol{X}$
$\boldsymbol{X}^m$	Masked load profile
$\boldsymbol{O}$	Output of BERT model
$\mathbf{T}$	Temperature profile

Abbreviations

BERT	Bidirectional Encoder Representations from Transformers
CV	Computer Vision
CVR	Conservation Voltage Reduction
DR	Demand Response
EGYE	Energy Error
FCE	Frequency Component Error
GAN	Generative Adversarial Network
LSTM	Long Short-Term Memory Network
MDS	Missing Data Segment
MPE	Mean Percentage Error
NLP	Natural Language Processing
PIN	Profile Inpainting Network
PoCP	Percentage of Closer Points
PKE	Peak Error
RMSE	Root Mean Square Error
SAE	Sparse Autoencoder
ViT	Vision Transformer
VLE	Valley Error

I. INTRODUCTION

THE restoration of missing data holds significant importance in power system analysis. In power system load profiles, two types of missing data exist. First, temporary equipment malfunctions or communication losses result in missing data, adversely affecting data quality. This impedes various data-related tasks such as load forecasting, load disaggregation, and anomaly detection. Second, demand response (DR) or conservation voltage reduction (CVR) baseline presents a unique case of missing data. For instance, utilities widely employ CVR for peak load reduction, where system voltage at the substation bus is decreased by 2-4% during a CVR event to achieve load reduction. However, the original load profile during a CVR event (the baseline), assuming no voltage reduction, remains unknown. Accurately estimating this baseline is crucial for load service providers to quantify the load reduction caused by CVR.

Current techniques for recovering missing data fall into two

This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office. Award Number: DE-EE0008770.

Yi Hu, Hyeonjin Kim, Kai Ye, and Ning Lu are with the Electrical & Computer Engineering Department, Future Renewable Energy Delivery and Management (FREEDM) Systems Center, North Carolina State University, Raleigh, NC 27606 USA. (emails: yhu28@ncsu.edu, hkim66@ncsu.edu, kye3@ncsu.edu, nlu2@ncsu.edu).

TABLE I
COMPARISON OF EXISTING POWER SYSTEM LOAD PROFILE INPAINTING METHODS

		Description	Advantages	Disadvantages
Model-based methods [1]-[5]		Use physical system models to simulate responses to external disturbances for restoring missing data segments.	Explainable as the models reflect the laws of physics.	Require accurate distribution system model.
Data-driven methods	Similarity-based [6]-[9]	Group load profiles by day type, weather conditions, and shape characteristics of load profiles. The missing data segments are restored by referencing to the data on the load profiles having the best similarity match.	Easy to implement and explainable.	Accuracy of the method dependent on selections of similarity metrics and weights.
	Regression-based (the benchmark method)	Use models including linear regression [10], Long Short Term Memory (LSTM) [11][12], Autoencoders [13][14], Gaussian Regression [15], Support Vector Regression (SVR) [16][17], etc. Or combine multiple regression models [18]-[21].	Provide transparent insights into the relationships between input and output. More efficient for small to medium-sized datasets.	Limited complexity and Limited contextual understanding. Require manual feature engineering.
	GAN-based [22]-[27]	Use Generative Adversarial Nets solve the missing data restoration problems in power system.	Discover underlying patterns in the data without explicit supervision.	Training instability, mode collapse, and hard to evaluate.
	Load-PIN [28] (the benchmark GAN-based method)	Combine Generative Adversarial Nets with Convolutional layers and multi-head self-attention blocks to improve accuracy.	More accurate than model-, similarity-, regression-, and other GAN based models.	Computationally expensive and require large amount of data. **Produce only 1 restoration candidate.**
	BERT-PIN (the proposed method)	Bidirectional Encoder can capture long-range dependencies though self-attention mechanisms.	The **most accurate** method and can produce **multiple restoration** candidates.	Computationally expensive and require large amount of data.

categories: model-based and data-driven methods. Table I provides a comprehensive overview and comparison of existing missing data restoration methods in power systems. Notably, Load-PIN [28], a GAN-based approach, outperforms model-based, similarity-based, regression-based, and other GAN-based methods in restoring missing data segments (MDSs) of fixed length.

However, existing inpainting techniques typically offer only a single solution. As illustrated in Fig. 1, we observed that MDSs can be analogized to missing words in sentences, thus equating the restoration process to recovering missing words in sentences or paragraphs. In the word completion task depicted in Fig. 1, if someone is fluent in French, "France" might emerge as the most probable choice. However, "Quebec" and "Cameroon" also represent viable alternatives, as they are regions where French is an official language. Similarly, when restoring an MDS, it's essential to provide multiple patching options, each with comparable likelihoods of being the best match.

Inspired by this observation, we use advanced Natural Language Processing (NLP) techniques to solve missing data restoration problem in power system domain. One of the advantages of employing NLP models is their capability to generate multiple alternatives for a missing word, each accompanied by a confidence level.

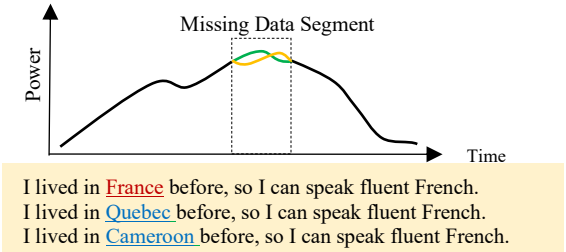

Fig. 1. An illustration of the load profile inpainting problem.

Since 2017, the Transformer model [29] and its variants, such as Bidirectional Encoder Representations from Transformers (BERT) [30], and Vision Transformer (ViT) [31], achieved remarkable success in NLP and computer vision (CV). They excel at handling sequential data and capturing extensive long-range dependencies by employing self-attention mechanisms that allow simultaneous consideration of all positions in a sequence. This introduction of self-attention empowers models to discern the importance of individual elements within input data, facilitating the understanding of intricate dependencies, relationships, and contextual information, ultimately leading to enhanced performance. It's important to clarify that the "Transformer" referred to in this context is an advanced machine learning model and not the transformer device used in power systems, to avoid any potential misunderstanding.

Therefore, we introduce the BERT-based Profile Inpainting Network (BERT-PIN), a flexible framework designed specifically for restoring multiple missing data segments within power system load profiles. To adapt the standard Transformer model structure for profile inpainting, we divide the load and temperature profiles into line segments, treating each segment as a word, the daily profile as a sentence, and the weekly/monthly load profile as a paragraph comprising multiple sentences. Furthermore, we integrate a top candidate selection process into BERT-PIN, allowing it to generate a series of probability distributions. Based on these distributions, users can produce multiple plausible imputed datasets, each reflecting varying confidence levels.

Although many machine-learning methods exist for restoring missing data, we selected the BERT-based model for three primary reasons. *First*, BERT's bidirectional nature makes it adept at capturing contextual information within time-series load profiles. *Second*, leveraging self-attention mechanisms enables BERT to effectively capture long-range dependencies,

enhancing its performance in processing sequential data. *Third*, BERT's architecture outputs a sequence of probability distributions, enabling the generation of multiple results.

Our contributions are summarized as following:

- First, to our knowledge, we are the first to introduce a BERT-based approach for power system load profile inpainting. BERT-PIN surpasses the state-of-the-art by approximately 5%-30%.
- Second, unlike existing methods, BERT-PIN can produce multiple data restoration candidates with varying confidence levels. This feature is particularly valuable when exploring all potential options is necessary to ensure algorithm robustness.
- Third, BERT-PIN can restore multiple MDSs within long-time windows. This flexibility allows BERT-PIN to be used in Demand Response baseline estimation, as well as various downstream tasks, such as load profile disaggregation [32] and super resolution [33].

The rest of the paper is organized as follows. Section II introduces the methodology, Section III introduces the simulation results in different cases, and Section IV concludes the paper.

II. Methodology

In this section, we first introduce the load profile inpainting problem formulation. Then, the BERT-PIN model architecture is illustrated in detail. Finally, the performance evaluation metrics are defined.

A. BERT-PIN Problem Formulation

Define $\boldsymbol{X}_m^{\square}$ as a MDS (the green segment in Fig. 2) in a time series load profile, $\boldsymbol{X} = [x_1, x_2, \dots, x_N]$, where N denotes the length of the time series.

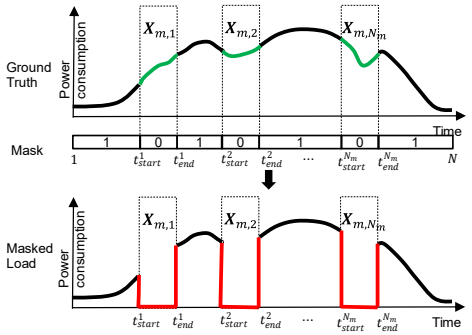

Fig. 2. Illustration of missing data segments and masking methods.

As shown in Fig. 2, if $\boldsymbol{X}$ contains N_m MDSs, i.e., $[\boldsymbol{X}_{m,1}^{\square}, \boldsymbol{X}_{m,2}^{\square}, \dots \boldsymbol{X}_{m,i}^{\square}, \dots, \boldsymbol{X}_{m,N_m}^{\square}]$, the objective of the inpainting problem is to find a set of model parameters, θ, to recover all MDSs using the non-missing data $\widetilde{\boldsymbol{X}}$ in $\boldsymbol{X}$ and the corresponding ambient temperature profile, $\boldsymbol{T} = [x_1, x_2, \dots, x_N]$ as inputs. So, the problem can be formulated as

$$[\widehat{\boldsymbol{X}}_{m,1}^{\square}, \widehat{\boldsymbol{X}}_{m,2}^{\square}, \dots, \widehat{\boldsymbol{X}}_{m,i}^{\square}, \dots, \widehat{\boldsymbol{X}}_{m,N_m}^{\square}] = f_\theta(T, \widetilde{\boldsymbol{X}}) \quad (1)$$

where $\widehat{\boldsymbol{X}}_{m,i}^{\square}$ is the i^{th} recovered MDS.

Next, we will introduce the design of the proposed BERT-PIN model to address the problem formulated above.

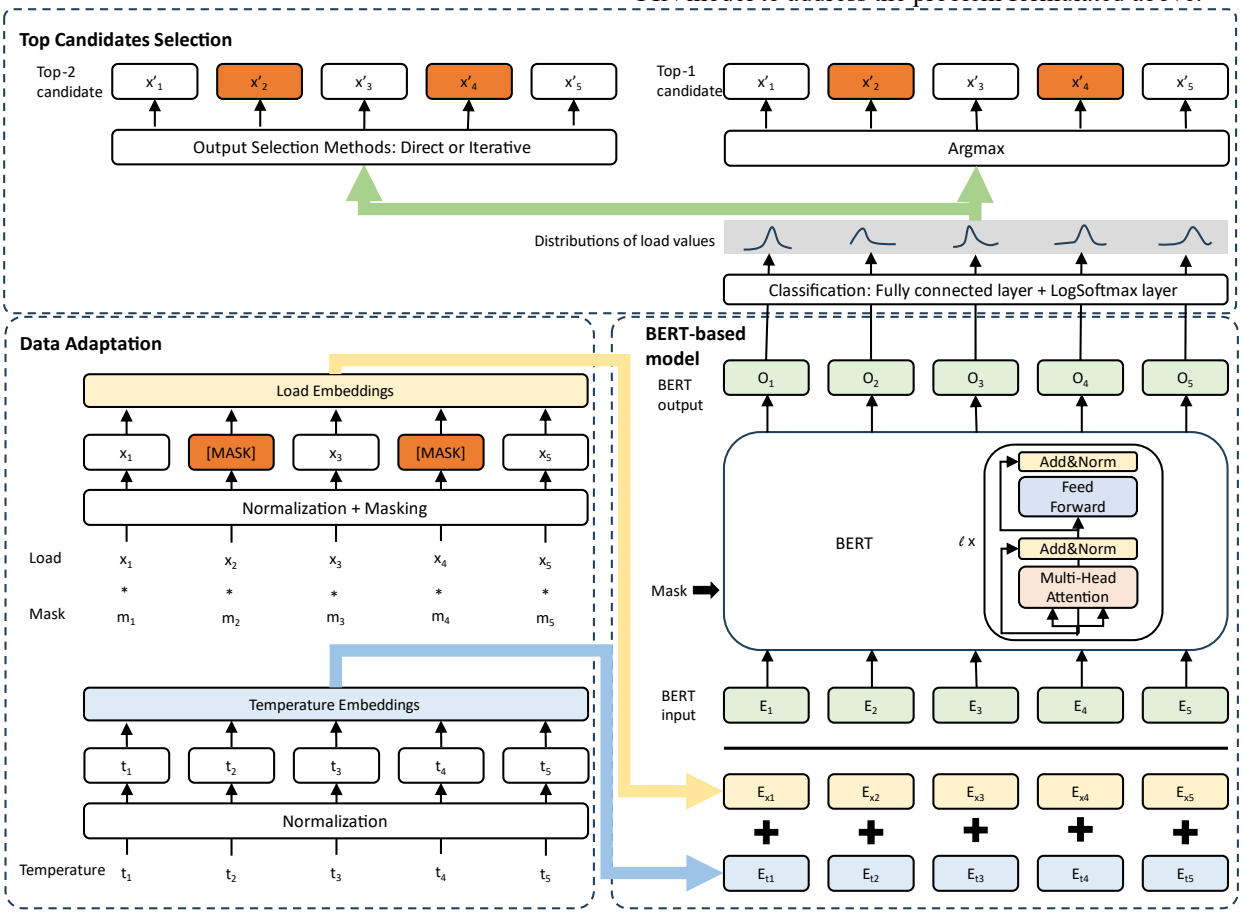

Fig. 3. An overview of the overall modeling framework. The illustration of the BERT was inspired by [29]. It's an example of sequence length $N = 5$.

B. BERT-PIN Model Architecture Overview

As illustrated in Fig. 3, the proposed BERT-PIN model comprises three fundamental processes: input data adaptation, BERT model for recovering MDSs, and top candidate selection.

Given the established efficacy of the BERT model structure in addressing NLP tasks, our intention is not to alter the original BERT model architecture. However, as the BERT model is initially designed for processing NLP problems, its inputs are a sequence of word tokens. Therefore, in the *Input Data Adaptation* process, we first align the load profile with its corresponding temperature profile. Then, we divide the two aligned time-series profiles into segments to generate the load and temperature embeddings, respectively. As depicted in Figs. 1 and 3, each of these segments resembles a missing word in a sentence, thus rendering the task of restoring missing data akin to recovering missing words within sentences or paragraphs.

During the *Top Candidates Selection* process, our objective is not to merely choose a single candidate with the highest likelihood. Instead, we aim to generate multiple candidates that meet predetermined confidence thresholds. This functionality enables BERT-PIN to generate an ensemble of patching options for a MDS. This feature proves especially critical when confronted with MDSs where several candidates display similar probabilities of being potential outcomes.

In the following sections, we will introduce those three main components in detail.

C. Data Adaptation

1) Preparation of Ground Truth Load Profiles

To prepare the training data, we collected smart meter data at 15-minute interval over a three-year period (2018-2020) from 8000 customers in North Carolina. Let P_i be the load profile for the i^{th} user containing N data points. Initially, we select a starting time (t_{start}) and calculate the end time (t_{end}) of the time series by $t_{end} = t_{start} + N - 1$. Next, we randomly draw N_{Agg} (ranges from 10 to 5000) load profiles (P) from the pool of 8000 load profiles and aggregate them into one load profile. This step is repeated for 200 times to ensure diversity in the training data. Finally, we normalize the aggregated load profile by its peak power (P_{MAX}) to create the ground truth load profile (X). This normalization ensures that X falls within the range of [0, 1]. This process can be summarized as

$$X = \frac{1}{P_{MAX}} \sum_{i=1}^{N_{Agg}} P_i\,(t_{start}: t_{end}) \tag{2}$$

$$P_{MAX} = \max \sum_{i=1}^{N_{Agg}} P_i\,(0: N) \tag{3}$$

2) Preparation of Load Profiles with Missing Data

To generate the time series data with N_m MDSs, we create a mask vector $\boldsymbol{M}$ for each $\boldsymbol{X}$, so that

$$\boldsymbol{M} = [m_i\ for\ i = 1{:}N], m_i = \begin{cases} 0, & missing\ data \\ 1, & otherwise \end{cases} \tag{4}$$

Then, the masked load profile, $\boldsymbol{X}^m$, can be represented as

$$\boldsymbol{X}^m = \boldsymbol{X} \cdot \boldsymbol{M} \tag{5}$$

Note that we set all missing data segments to be 0 kW because all the aggregated power values are greater than 0 kW, making it a unique value to be distinguishable.

3) Input Data Adaptation Layer

To align the BERT-PIN inputs with the BERT required format, we map the values in $\boldsymbol{X}^m$ to integers between 0 to 200. Note that we chose 200 because it strikes a balance between the model's size and resolution. In our dataset with 2000 aggregation level, the power range is [210, 1751] kW, as shown in Fig. 4. So, the mapping provides a resolution of 8.755 kW.

The mapped load profile is embedded into a $N \times 200$ matrix, represented by the load embeddings (yellow boxes) in Fig. 3. This data adaptation process allows the model to generate a probability distribution for each data point, making it possible to generate an ensemble of candidates. This transforms the missing data restoration problem, which is usually a regression problem, into a classification problem.

To address the influence of temperature on load [34], we include the normalized ambient temperature profile data, designated as T, as an additional modality input to assist in the recovery of missing load data. T is subjected to normalization based on the highest and lowest annual temperatures, ensuring that the normalized temperature values also range from 0 to 1. Then, T is rescaled to the [0, 200] range using the same approach employed for load embedding.

Lastly, we combine $\boldsymbol{X}^m$ and T embeddings together by element-wise addition to obtain the final input matrix, the dimension of which is $N \times 200$. As an illustration, we show the input data adaptation process when $N = 5$ in Fig. 3.

D. BERT Model

The combined embeddings of load and temperature can be directly fed into the BERT model using its original model architecture introduced in [30]. Represent a data sequence as $\boldsymbol{D} = \{\{\boldsymbol{k}_1, \boldsymbol{v}_1\}, \ldots \{\boldsymbol{k}_N, \boldsymbol{v}_N\}\}$, where $\boldsymbol{k}$ and $\boldsymbol{v}$ are N tuples of keys and values, respectively. For a query $\boldsymbol{q}$, the attention [35] over $\boldsymbol{D}$ in the BERT model is formulated as

$$Attention(\boldsymbol{q}, \boldsymbol{D}) = \sum_{i=1}^{N} \alpha(\boldsymbol{q}, \boldsymbol{k}_i)\, \boldsymbol{v}_i \tag{6}$$

where $\alpha(\boldsymbol{q}, \boldsymbol{k}_i) \in \mathbb{R}\ (i = 1, \ldots, N)$ are scalar values calculated through the dot product of the query vector and each key vector. These weights, referred to as attention scores, are normalized to ensure a sum of 1. This calculation is performed for every element in the input sequence. The output vectors are combined to form the final output of the model.

When processing sequential data, using self-attention can selectively attend to different parts of the input sequence based on their relevance to the given query vector. Thus, the BERT model can effectively capture long-range dependencies to form a context for the sequence. This significantly facilitates the recovery of the missing data because the context reflects the relevance of all known data points from all modalities (e.g., load and temperature) with the missing data points in those time-series profiles.

In the training, we use $\hat{\boldsymbol{X}}^{\boldsymbol{1}}$ to calculate the cross-entropy

losses [36] as

$$CrossEntropy = -\frac{1}{N}\sum_{o=1}^{N}\sum_{c=1}^{C} x_{o,c}\log\left(d_{o,c}\right) \quad (7)$$

where C is the number of classes, $x_{o,c}$ is the truth label denoting the power consumption value for observation o, and $d_{o,c}$ is the predicted probability observation o belonging to class c. N is the length of the sequence.

Note that the objective of BERT-PIN is to effectively restore the missing segments in a load profile, so we need to train the network to place more focus on restoring the MDSs. Thus, we construct the loss function as

$$Loss = (1-\lambda) * CrossEntropy\left(\boldsymbol{X}, \widehat{\boldsymbol{X}}_{\square}^{\boldsymbol{1}}\right) + \lambda * CrossEntropy(\boldsymbol{X}_m, \widehat{\boldsymbol{X}}_{\boldsymbol{m}}^{\boldsymbol{1}}) \quad (8)$$

where λ is a hyper parameter for balancing between the global and local losses. Thus, $(1-\lambda) * CrossEntropy\left(\boldsymbol{X}, \widehat{\boldsymbol{X}}_{\square}^{\boldsymbol{1}}\right)$ represents the global loss for assessing how well the whole load profile can be restored, and $\lambda * CrossEntropy(\boldsymbol{X}_m, \widehat{\boldsymbol{X}}_{\boldsymbol{m}}^{\boldsymbol{1}})$ represents the local loss for assessing how well the MDSs are restored.

E. Top Candidates Selection Layer

As illustrated in Fig. 3, we add a *Top Candidates Selection* process. Define the output of the BERT model, $\boldsymbol{O}$, as:

$$\boldsymbol{O} = BERT(\boldsymbol{X}^m, \boldsymbol{T}) \quad (9)$$

where $\boldsymbol{O}$ is an $N \times 200$ matrix.

We feed $\boldsymbol{O}$ into a classification layer, which comprises a fully connected layer followed by a SoftMax layer, to obtain an $N \times 200$ probability distribution matrix, $\boldsymbol{D}$. Note that the i^{th} column of $\boldsymbol{D}$ represents the probability distribution function (PDF) of the value of the i^{th} data point falling within the range of 1 to 200.

$$\boldsymbol{D} = Classification(\boldsymbol{O}) \quad (10)$$

Using $\boldsymbol{D}$, we can generate an ensemble of curves for patching an MDS rather than outputting just a single curve with the highest likelihood.

The conventional method for restoring the MDS from $\boldsymbol{D}$ is to use an *argmax* layer as illustrated as the orange boxes in Fig. 3 by

$$\widehat{\boldsymbol{X}}_{\square}^{\boldsymbol{1}} = argmax(\boldsymbol{D}) \quad (11)$$

where $\widehat{\boldsymbol{X}}_{\square}^{\boldsymbol{1}}$ is considered as the *top-1 candidate*.

Nevertheless, it is often necessary to explore the top-2 or even the top-3 candidates as potential patching options to enhance the inclusion. As depicted in Fig. 1, a single blank in a sentence can have multiple possible words (i.e., "France", "Quebec" and "Cameroon") that fit the context of the original sentence. When used for nationality identification, it becomes imperative to supply a list of regions where French is spoken as an official language, ranked in order of population size.

Similarly, there may exist multiple plausible curves for patching an MDS. The *top-1* method depicted in (11) selects the candidate with the highest probability for each missing data point. If the PDF of the missing data has a higher peak or are more narrowly concentrated around a particular value (see the red PDF in Fig. 4), this method may have a higher accuracy in selecting the best candidate. However, when the PDF curve is a somewhat flattened one (see the green distribution in Fig. 4), selecting only the best candidate will greatly limit the inclusion of the original missing data point.

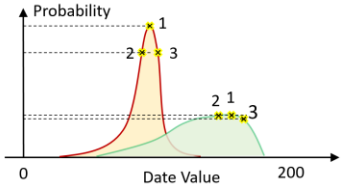

Fig. 4. Illustration of the probability distribution in *D*.

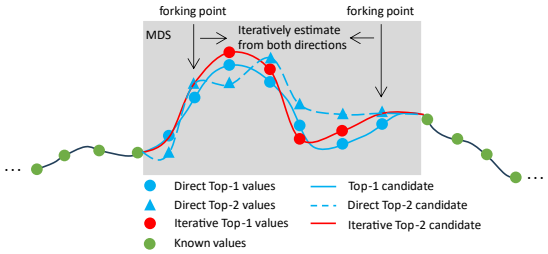

Fig. 5. Illustration of the top-2 candidate selection process.

Therefore, we extend the top-1 method to the top-2 method, by comparing two selection approaches. Note that this approach can be readily extended to top-X candidate selection by repeatedly applying the same selection criterion.

Method 1: The simplest approach involves selecting the candidate with the second-highest probability from each probability distribution to create an additional set of profiles. As illustrated in Fig. 5, the solid blue curve connecting the blue circles represents the curve generated by the top-1 candidates, while the dashed blue curve connecting the blue triangles represents the patching curve generated by directly selecting the top-2 candidates using Method 1.

However, Method 1 suffers from a significant drawback: it ignores the autocorrelation among adjacent data points. In practical terms, selecting the second candidate may lead to shifts in the probability density functions (PDFs) of subsequent missing data points.

Method 2: To overcome the inherent limitation of Method 1, we propose an iterative selection approach centered on the initial identification of "fork points". By using the identified fork points as reference pivots, we can create the top-2 candidate curve accounting for the autocorrelation among the subsequent data points.

Define the "right-side" (or "left-side") fork point as the first point, counting from the rightmost (or leftmost) side of the MSD, where the probability difference between the top-1 and top-2 candidates is less than e. As depicted in Fig. 5, after identifying the fork points, we can start an iterative process to estimate the top-1 values for the remaining missing data points, one at a time, by shifting the load profile either forward or backward. This method ensures that the shifted load profile maintains the same masking position as the original one, effectively capturing the shift in PDFs for subsequent data

points.

The resultant top-2 curve is shown by the red continues line in Fig. 5 and the detailed algorithm description is depicted in Algorithm 1.

Algorithm 1 *Iterative Top Candidates Selection*.
Given the output of BERT-PIN $\boldsymbol{D}$, find the fork points located on both the left and right sides of the top-1 curve. Next, replace the top-1 values at the fork points with the top-2 values. Then, iteratively generate the remaining missing data using top-1 values. Note that e is the threshold for selecting the fork points.

Let $l = (t_{end} - t_{start})/2$
for $t = t_{start}: t_{start} + l$ **do**
    # find the left fork point.
    if $top1(\boldsymbol{D}_t) - top2(\boldsymbol{D}_t) < e$ **do**
        $fork_{left} = t$
        $x_t^{masked} =$ index of $top2(\boldsymbol{D}_t)$
        break
    else
        $x_t^{masked} =$ index of $top1(\boldsymbol{D}_t)$
end for
for $t = t_{end}: t_{end} - l$ **do**
    # find the right fork point.
    if $top1(\boldsymbol{D}_t) - top2(\boldsymbol{D}_t) < e$ **do**
        $fork_{right} = t$
        $x_t^{masked} =$ index of $top2(\boldsymbol{D}_t)$
        Break
    else
        $x_t^{masked} =$ index of $top1(\boldsymbol{D}_t)$
end for

for $k = fork_{left} - t_{start}: l$ **do**
- shift the daily profile by k steps to the left.
- feed shifted data into the BERT-PIN model.
  $\boldsymbol{D}' = BERT(\boldsymbol{X}^{shifted}, \boldsymbol{T}^{\boldsymbol{shifted}})$
- update the first unknown value in $\boldsymbol{X}^{masked}$
  $x_{t_{start}+k}^{masked} =$ index of $top1(\boldsymbol{D}'_{t_{start}})$

end for

for $k = t_{end} - fork_{right}: l$ **do**
- shift the daily profile by k steps to the right.
- feed shifted data into the BERT-PIN model.
  $\boldsymbol{D}' = BERT(\boldsymbol{X}^{shifted}, \boldsymbol{T}^{\boldsymbol{shifted}})$
- update the last unknown value in $\boldsymbol{X}^{masked}$
  $x_{t_{end}-k}^{masked} =$ index of $top1(\boldsymbol{D}'_{t_{end}})$

end for

Use the final $\boldsymbol{X}^{masked}$ as the restored *top-2* load profile, $\hat{\boldsymbol{X}}$.

F. Performance Metrics

The performance metrics used for evaluating the accuracy of the restored data segments are calculated as

$$MPE = \frac{1}{K}\sum_{t=1}^{K}\frac{|\hat{x}_t^m - x_t^m|}{x_t^{event}} \quad (12)$$

$$RMSE = \sqrt{\frac{1}{K}\sum_{t=1}^{K}(\hat{x}_t^m - x_t^m)^2} \quad (13)$$

$$PKE = \frac{|\hat{x}^{MAX} - x^{MAX}|}{x^{MAX}} \quad (14)$$

$$VLE = \frac{|\hat{x}^{MIN} - x^{MIN}|}{x^{MIN}} \quad (15)$$

$$EGYE = \frac{|\sum_{t=1}^{K}\hat{x}_t^m - \sum_{t=1}^{K}x_t^m|}{\sum_{t=1}^{K}x_t^{event}} \quad (16)$$

$$FCE = \frac{1}{K}\sum_{t=1}^{K}\frac{|FFT(\hat{x}_t^m) - FFT(x_t^m)|}{FFT(x_t^m)} \quad (17)$$

where K is the total number of data points in the MDS, $\hat{x}$ represents the restored data segment, x^{MAX} and x^{MIN} are the maximum and minimum power values in the MDS, respectively, and FFT stands for Fast Fourier Transform. These indices offer insights into different aspects of errors, including point-to-point errors, inaccuracies in peak and valley points, discrepancies in total energy consumption, and errors within the frequency domain components.

III. Simulation Results

In this section, we evaluate BERT-PIN's capability to recover varying numbers of MDSs in different scenarios and compare it with three benchmark methods: LSTM [11], SAE [13], and Load-PIN [28]. The evaluation encompasses various aggregation levels and training data volumes. Additionally, we assess the efficacy of the top-1 and top-2 candidate methods in CVR baseline estimation to highlight the benefits of having several restoration alternatives at our disposal.

A. Data Preparation

The load profiles used in this study consists of 15-minute resolution smart meter data obtained from 8,000 users (including residential, commercial, and industrial users) in North Carolina between 2018 and 2020. The corresponding temperature data is downloaded from the National Oceanic and Atmospheric Administration (NOAA) [37] website and is used as a second modality input.

We randomly select a group of users from the pool of 8,000 and combine their data to form aggregated load profiles. The selected group size ranges from 10 to 5000, to show the model's capability to deal with different scenarios. An example of the distribution of load values and daily peak values in 2000 user aggregation level is shown in Fig. 6. Moreover, to increase the data diversity, the selection is repeated for multiple times. These aggregated load profiles are aligned with the temperature profile and then normalized based on the annual load and temperature peaks, respectively. Then, we partition the profiles into either daily (96 data points) or weekly profiles (672 data points). Each missing load data segment is consistently set at 4 hours (16 data points), a choice guided by the observation that around 70% of missing load data segments in actual utility data are less than 4 hours in duration. It's important to note that there are no missing data segments in the temperature profile. The dataset is divided into an 80% training set and a 20% testing set.

The hyperparameters and simulation platform information are summarized in Appendix Table A.I for reference. In the next sections, we will show the simulation results.

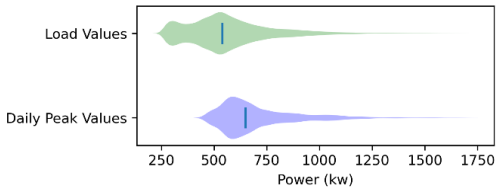

Fig. 6. Distribution of the 2000 users aggregated load values and daily peak values.

B. Restoration of One MDS

In this base case, we aggregate randomly selected 2000 users for 200 times. Our objective is to evaluate the performance of BERT-PIN when restoring a single 4-hour MDS. The daily load profile, with a 4-hour gap, together with the temperature data of the same day is used as input. And BERT-PIN restores the MDS. Several restored daily load profiles are shown in Fig. 7, while the corresponding performance metrics are depicted in Fig. 8 and summarized in Table II. It is evident from the outcomes that BERT-PIN (represented by the red lines) exhibits the smallest medians and ranges in this case. Furthermore, Table II provides the mean errors for each distribution shown in Fig. 7. The results clearly indicate that BERT-PIN achieves the lowest errors and outperforms the benchmark methods by 5%-27% across all six indexes. This demonstrates that our model consistently outperforms the existing approaches by higher missing data patching accuracy.

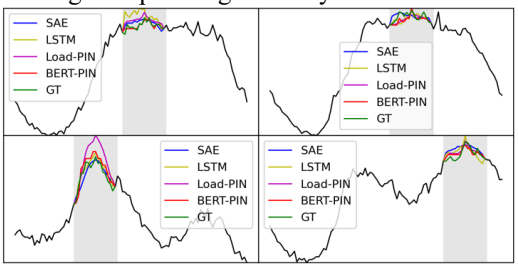

Fig. 7. Examples of missing data restoration with central-mask using different models: SAE (blue), LSTM (yellow), Load-PIN (magenta), BERT-PIN (red), and ground truth (green).

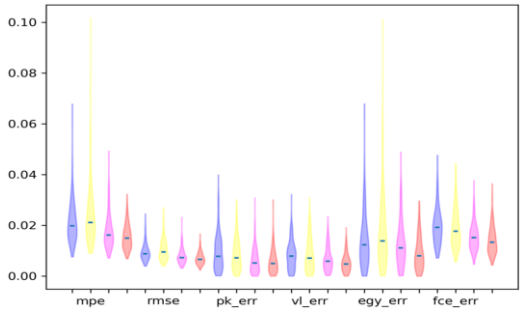

Fig. 8. Error distributions of different models: SAE (blue), LSTM (yellow), Load-PIN (magenta), BERT-PIN (red).

TABLE II
ERRORS OF SINGLE MDS INPAINTING

	SAE (%)	LSTM (%)	Load-PIN (%)	**BERT-PIN (%)**	**Improvement**
MPE	2.231	2.414	1.670	**1.523**	8.80%
RMSE	1.035	1.112	0.7951	**0.7404**	6.88%
PKE	1.065	0.8491	0.6183	**0.5130**	17.10%
VLE	0.8687	0.8991	0.6185	**0.5870**	5.09%
EGYE	1.525	1.762	1.165	**0.8410**	27.81%
FCE	2.138	2.046	1.615	**1.509**	6.56%

C. Expand the Simulation to Different Aggregation Levels and Different Data Sizes

1) Accuracy vs. Aggregation Level

In the preceding section, we maintained a fixed aggregation level of 2000. However, we now extend the aggregation level range from 10 to 5000 to demonstrate BERT-PIN's ability to handle diverse datasets.

The findings are illustrated in Fig. 9. As the aggregation level rises, errors diminish. A lower aggregation level typically implies greater randomness in the aggregated load profile, posing challenges for MDS restoration by the models. Moreover, irrespective of the aggregation level, BERT-PIN consistently demonstrates higher accuracy compared to the other three methods. This showcased BERT-PIN's effectiveness across various load aggregation levels.

2) Accuracy vs. Data Size

In this section, we assess how the size of the dataset affects training precision, taking the aggregation of 2000 loads as a case study. The model is trained with datasets ranging from 10 to 400 sets of load profiles. As illustrated in Fig. 10, a clear reduction in error rates is observed as the dataset size expands from 10 to 50 sets of load profiles, after which the improvements plateau. This suggests that aggregating data from 2000 randomly selected users and replicating this aggregation 200 times to generate 200 aggregated load profiles yields adequate outcomes. Expanding the dataset beyond this point does not significantly enhance model performance but results in disproportionately extended training durations.

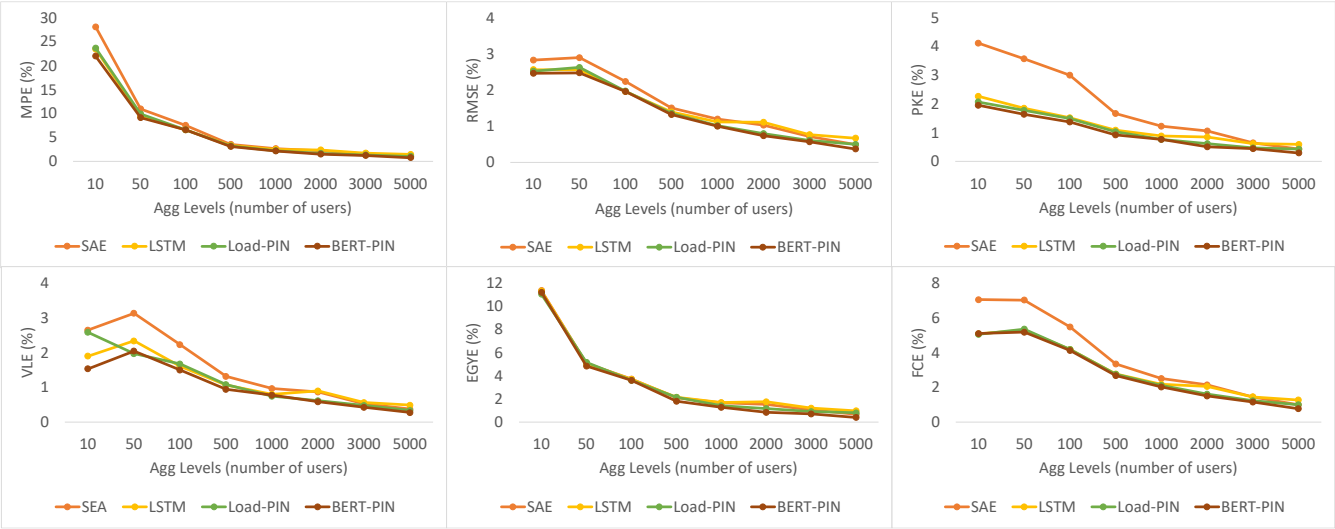

Fig. 9. Errors with different aggregation levels.

TABLE III
ERRORS OF TOP-1 CANDIDATE, TOP-2 CANDIDATES AND COMBINED OUTPUTS (%)

	BERT-PIN	BERT-PIN_2	BERT-PIN_2i						**Combine**
			e=0.8	e=0.5	e=0.3	e=0.1	e=0.05	e=0.02	
MPE	**1.523**	**1.744**	2.556	2.433	2.407	2.138	1.87	1.761	1.211
RMSE	**0.7404**	0.9144	1.317	1.173	1.211	1.071	0.896	**0.899**	0.577
PKE	**0.5130**	**0.5917**	1.044	0.939	0.927	0.871	0.665	0.663	0.426
VLE	**0.5870**	0.9260	0.827	0.617	0.817	0.573	**0.543**	0.669	0.407
EGYE	**0.8410**	0.9582	1.618	1.441	1.412	1.301	1.043	**0.986**	0.633
FCE	**1.509**	1.942	2.447	2.18	2.273	2.007	**1.727**	1.78	1.209
PoCP	-	45.12%	23.88%	24.40%	22.94%	17.29%	12.25%	**6.29%**	-

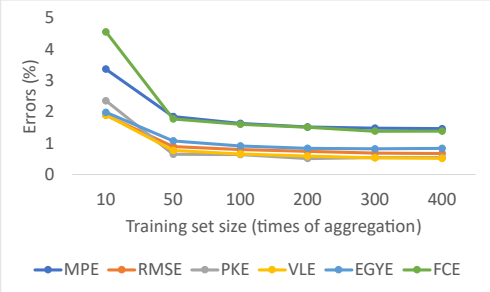

Fig. 10. BERT-PIN errors with different training data sizes.

D. Top-2 Candidate Selection

As described in section II, BERT-PIN has the capability to generate multiple restored load profiles using *Top Candidate Selection* methods. In this section, we compare the two Selection methods discussed in section II.D, taking the selection of the top-2 candidates as an illustrative example. It's important to note that this method can be employed iteratively to choose the top-X candidates.

The result is presented in Fig. 11. BERT-PIN is the default *top-1* candidate where the candidate with the highest probability value is selected. BERT-PIN_2 results are obtained using Method 1, where candidates having the *second highest* probability is selected. BERT-PIN_2i results are obtained using method 2, where an iterative method is used to select a fork point, based on which, subsequent restoration candidates are selected.

The errors of *top-1* candidate and *top-2* candidate with different parameters are computed and presented in Table III.

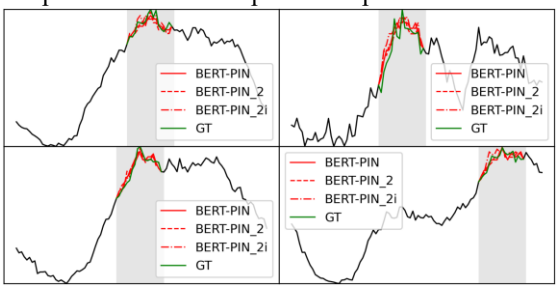

Fig. 11. Examples of top-1 and top-2 restored MDSs.

Given that the process can be repeated for generating an ensemble of candidates for MDS restoration, we proceed to assess the quality of the top-2 results by calculating the "percentage-of-closer-points" (PoCP). This metric signifies the percentage of estimated points that are closer to the ground truth when compared to those estimated using the top-1 method, indicating the potential expansion of inclusion through the incorporation of the top-2 restored MDS.

From the results shown in Table III, we made the following observations:

- As expected, the result accuracy of the *top-1* method is higher than the *top-2* method.
- Nevertheless, the *top-2* method can generate points that are closer to the ground truth. This is because the BERT output is generated based on probabilities, and the second-best candidate could also be close to the actual missing data, just with reduced likelihood.
- Compared with the direct *top-2* candidate selection method, iterative *top-2* candidate selection method has lower PoCP. However, since the points selected independently lack autocorrelation among adjacent points, we consider the results to be less valuable for time-series MDS.
- Among all iterative top-2 methods, when $e = 0.5$ (column highlighted in yellow), the PoCP is the highest. This demonstrates that the fork point selection influences the range of inclusion. When e becomes smaller, the iterative top candidate selection algorithm tends to identify a fork point later. Consequently, the top-2 candidate curve closely resembles the top-1 candidate, leading to identification of fewer "better" points.
- When we merge the top-1 and top-2 candidate by selecting the best value for each timestamp (column highlighted in gray), the errors are even smaller than the top-1 result. This demonstrates the benefit of incorporating both candidates when performing downstream tasks. This advantage is reflected in the wider range of data encompassed because, with both curves, we capture data points closer to the original missing data points.

Although the top-1 candidate delivers the highest accuracy, integrating either direct or iterative approaches with top-2 candidates broadens the inclusiveness of missing data estimation. Combining the top-1 and top-2 candidates enhances the robustness of our simulations beyond what's achievable with only the top-1 candidate. This approach also showcases BERT-PIN's capability to generate an ensemble of data segments for more comprehensive missing data restoration.

E. Restoration of Multiple MDSs

In this section, we showcase the performance of BERT-PIN for restoring multiple MDSs within a weekly load profile. This use case is essential for Conservation Voltage Reduction (CVR)

baseline estimation. CVR is a frequently adopted strategy among utility companies to manage peak loads. During CVR events, the voltage on a distribution feeder is deliberately decreased by 2-4% over a period ranging from 1 to 4 hours. CVR can be executed on multiple days in a hot summer week or a cold winter week. As depicted in Fig. 12, accurately assessing the actual load reduction achieved through CVR often involves estimating the baseline energy consumption (indicated by the red lines) that would occur in the absence of CVR implementation.

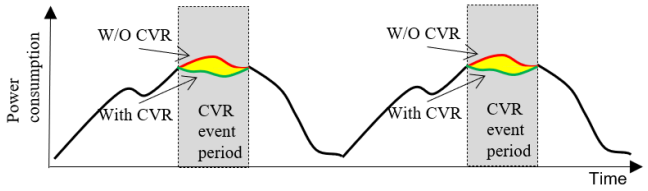

Fig. 12. An illustration of the CVR baseline estimation

To train BERT-PIN for CVR baseline estimation, we apply 4-hour masks targeting the peak load periods within weekly load profiles. The number of masks is randomly selected between 1 and 7, thereby representing different numbers of CVR event days. The hyperparameters of the BERT-PIN model remain consistent with those employed in the single-event scenario. As can be seen from Fig. 13 and Table IV, BERT-PIN outperforms all existing methods by a large margin.

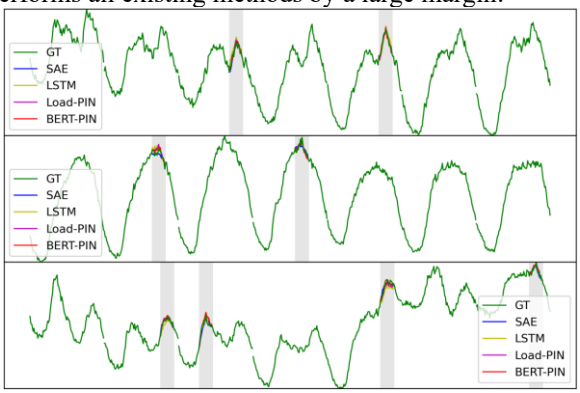

Fig. 13. Missing data restoration examples: SAE (blue), LSTM (yellow), Load-PIN (magenta), BERT-PIN (red), and ground truth (green).

TABLE IV
ERRORS OF MULTIPLE MDSS INPAINTING

	SAE (%)	LSTM (%)	Load-PIN (%)	**BERT-PIN (%)**	**Improvement**
MPE	7.615	6.389	5.168	**4.837**	**6.40%**
RMSE	3.310	2.848	2.301	**2.221**	**3.48%**
PKE	3.362	2.193	1.769	**1.746**	**1.30%**
VLE	2.697	2.407	1.691	**1.627**	**3.78%**
EGYE	4.874	3.822	3.257	**2.699**	**17.13%**
FCE	6.828	5.601	4.808	**4.501**	**6.39%**

This section demonstrates that BERT-PIN can effectively restore multiple missing data segments and accurately estimate the baseline of demand response programs with CVR baseline estimation serving as a practical case study.

IV. Conclusion

In this study, we introduced a cutting-edge framework called BERT-PIN (Bidirectional Encoder Representations from Transformers-based Profile Inpainting Network), leveraging the advanced capabilities of the BERT model to restore multiple missing data segments within load profiles. Our contributions can be summarized in three main aspects. First, to the best of our knowledge, we are the first to introduce a BERT-based approach for power system load profile inpainting, leading to higher accuracy. Second, unlike existing methods, BERT-PIN can produce multiple data restoration candidates with varying levels of confidence, offering enhanced flexibility in data recovery. Third, BERT-PIN can restore multiple missing data segments (MDSs) within very long-time windows.

We tested our BERT-PIN model on different datasets from North Carolina, covering various scenarios like different levels of data aggregation and dataset sizes. The results show that BERT-PIN outperforming other methods by 5%-30% in accuracy for filling in both single and multiple gaps in the data. Moreover, using a combination of the top-1 and top-2 predictions allows us to predict missing data more comprehensively. Overall, BERT-PIN proved effective in specific tasks like filling in weekly data gaps and demand response baseline estimation, outperforming traditional methods in accuracy.

In our future research endeavors, we intend to apply the BERT model, already pre-trained, to a variety of downstream tasks such as Super Resolution, Load Disaggregation, Anomaly Detection, and Energy Forecasting. Our goal is to investigate how the attention mechanism and contextual capabilities of BERT can enhance both the precision and reliability of these specific tasks.

Appendix

To facilitate easy replication, we have made our source code accessible at: https://github.com/hughwln/BERT-PIN_public

TABLE A.I
SYSTEM PARAMETERS SELECTED FOR BERT-PIN

Parameters	**Values**
Learning rate	1e-4
Local loss weight λ	0.8
Batch size	16
Training epochs	100
Number of heads	2
Number of transformer layers ℓ	2
Training platform	NVIDIA GeForce RTX 4090
Training time	6 hours
Programing environment	Python 3.8 + PyTorch 2.1.0

References

[1] I. Visconti, D. A. Lima and J. V . Milanović. "Comprehensive analysis of conservation voltage reduction: A real casestudy." 2019 IEEE Milan PowerTech. IEEE, 2019.

[2] Y. Zhang, S. Ren, Z. Dong, Y. Xu, K. Meng, and Y. Zheng, "Optimal placement of battery energy storage indistribution networks considering conservation voltage reduction and stochastic load composition." IET Generation, Transmission & Distribution, vol. 11, no. 15, pp. 3862-3870, 2017.

[3] Z. Wang, and J. Wang. "Time-varying stochastic assessment of conservation voltage reduction based on load modeling." IEEE Transactions on Power Systems, vol. 29, no. 5, pp. 2321-2328, 2014.

[4] M. Diaz-Aguiló, J. Sandraz, R. Macwan, F. Leon, D. Czarkowski, C. Comack, and D. Wang, "Field-validated load model for the analysis of CVR in distribution secondary networks: Energy conservation." IEEE Transactions on Power Delivery, vol. 28, no. 4, pp. 2428-2436, 2013.

[5] K.P. Schneider, F.K. Tuffner, J.C. Fuller, and R. Singh, "Evaluation of conservation voltage reduction (CVR) on a national level." Pacific Northwest National Lab (PNNL), Richland, WA (United States), no. PNNL-19596, 2010.

[6] K. Coughlin, M. Piette, C. Goldman, and S. Kiliccote, "Estimating demand response load impacts:Evaluation of baselineload models for non-residential buildings in california." Lawrence Berkeley National Lab (LBNL), Berkeley, CA (United States), no. LBNL-63728, 2008.

[7] B. Xiang, K. Li, X. Ge, F. Wang, J. Lai, and P. Dehghanian, "Smart Households' Available Aggregated Capacity Day-ahead Forecast Model for Load Aggregators under Incentive-based Demand Response Program." 2019 IEEE Industry Applications Society Annual Meeting. IEEE, 2019.

[8] T.K. Wijaya, M. Vasirani, and K. Aberer. "When bias matters: An economic assessment of demand response baselines for residential customers." IEEE Transactions on Smart Grid, vol. 5, no. 4, pp. 1755-1763, 2014.

[9] H. P. Lee, L. Song, Y. Li, N. Lu, D. Wu, PJ Rehm, M. Makdad, E. Miller, "An Iterative Bidirectional Gradient Boosting Algorithm for CVR Baseline Estimation" 2023 IEEE Power & Energy Society General Meeting (PESGM), 2023.

[10] S. Matsukawa, C. Ninagawa, J. Morikawa, T. Inaba, and S. Kondo, "Stable segment method for multiple linear regression on baseline estimation for smart grid fast automated demand response." 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia). IEEE, 2019.

[11] J. Oyedokun, S. Bu, Z. Han, and X. Liu, "Customer baseline load estimation for incentive-based demand response using long short-term memory recurrent neural network." 2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe). IEEE, 2019.

[12] C, Bülte, M. Kleinebrahm, H.Ü. Yilmaz, and J. Gómez-Romero, "Multivariate time series imputation for energy data using neural networks." Energy and AI, vol. 13, pp.100239, July 2023.

[13] Y. Chen, C. Chen, X. Zhang, M. Cui, F. Li, X. Wang, and S. Yin, "Privacy-Preserving Baseline Load Reconstruction for Residential Demand Response Considering Distributed Energy Resources." IEEE Transactions on Industrial Informatics, vol. 18, no. 5, pp. 3541-3550, 2022.

[14] A. Liguori, R. Markovic, M. Ferrando, J. Frisch, R. Causone, and C. Treeck, "Augmenting energy time-series for data-efficient imputation of missing values." Applied Energy, vol. 334, pp. 120701, March 2023.

[15] Y. Weng, Y. Jiafan, and R. Rajagopal. "Probabilistic baseline estimation based on load patterns for better residential customer rewards." International Journal of Electrical Power & Energy Systems, no. 100, pp. 508-516, 2018.

[16] Y. Chen, P. Xu, Y. Chu, W. Li, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the Support Vector Regression (SVR) model to calculate the demand response baseline for office buildings." Applied Energy, no. 195, pp. 659-670, 2017.

[17] Z. Wang, M. Begovic, and J.Wang. "Analysis of conservation voltage reduction effects based on multistage SVR and stochastic process." IEEE Transactions on Smart Grid, vol. 5, no. 1, pp, 431-439, 2013.

[18] M. Sun, Y. Wang, F. Teng, Y. Ye, G. Strbac, and C. Kang, et al. "Clustering-based residential baseline estimation: A probabilistic perspective." IEEE Transactions on Smart Grid, vol. 10, no. 6, pp. 6014-6028, 2019.

[19] Y. Zhang, Q. Ai, and Z. Li. "Improving aggregated baseline load estimation by Gaussian mixture model." Energy Reports, vol. 6, pp. 1221-1225, 2020.

[20] X. Ge, F. Xu, Y. Wang, H. Li, F. Wang, J. Hu, X. Lu, and B. Chen, "Spatio-Temporal Two-Dimensions Data Based Customer Baseline Load Estimation Approach Using LASSO Regression." IEEE Transactions on Industry Applications, vol. 58, no. 3, pp. 3112-3122, 2022.

[21] J. Chen, J. Yuan, W. Chen, A. Zeb, M. Suzauddola, and Y. A. Nanehkaran. "Research on Interpolation Method for Missing Electricity Consumption Data." Computers, Materials & Continua, vol, 78, no.2, 2024.

[22] J. Yoon, J. Jordon, and M. Schaar. "Gain: Missing data imputation using generative adversarial nets." International conference on machine learning. PMLR, 2018.

[23] Y. Luo, X. Cai, Y. Zhang, and J. Xu. "Multivariate time series imputation with generative adversarial networks." Advances in neural information processing systems 31, 2018.

[24] K. Zhang, X. Dou, and X. Xiao. "Grid Defect Data Completion Based on Generative Adversarial Imputation Nets." 2021 IEEE Sustainable Power and Energy Conference (iSPEC). IEEE, 2021.

[25] W. Zhang, Y. Luo, Y. Zhang, and D. Srinivasan. "SolarGAN: Multivariate solar data imputation using generative adversarial network." IEEE Transactions on Sustainable Energy, vol. 12, no. 1, pp. 743-746, June 2020.

[26] X. Hu, Z. Zhan, D. Ma and S. Zhang, "Spatiotemporal generative adversarial imputation networks: An approach to address missing data for wind turbines", IEEE Trans. Instrum. Meas., vol. 72, pp. 1-8, 2023.

[27] S. Ma, Z.-S. Xu and T. Sun, "Parallel generative adversarial imputation network for multivariate missing time-series reconstruction and its application to aeroengines", IEEE Trans. Instrum. Meas., vol. 72, pp. 1-16, 2023.

[28] Y. Li, L. Song, Y. Hu, H.P. Lee, D. Wu, PJ Rehm, and N. Lu, "Load Profile Inpainting for Missing Load Data Restoration and Baseline Estimation.", IEEE Transactions on Smart Grid, vol. 15, no. 2, pp. 2251-2260, Mar. 2024, doi: 10.1109/TSG.2023.3293188.

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need." In NIPS, 2017.

[30] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding." In NAACL, 2019.

[31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[32] K. Ye, H. Kim, Y. Hu, N. Lu, D. Wu, and P. Rehm, "A Modified Sequence-to-point HVAC Load Disaggregation Algorithm," 2023 IEEE Power & Energy Society General Meeting (PESGM), Orlando, FL, USA, 2023, pp. 1-5, doi: 10.1109/PESGM52003.2023.10252553.

[33] L. Song, Y. Li and N. Lu, "ProfileSR-GAN: A GAN Based Super-Resolution Method for Generating High-Resolution Load Profiles," in IEEE Transactions on Smart Grid, vol. 13, no. 4, pp. 3278-3289, July 2022, doi: 10.1109/TSG.2022.3158235.

[34] S. Haben, G. Giasemidis, F. Ziel, and S. Arora, "Short term load forecasting and the effect of temperature at the low voltage level," International Journal of Forecasting, vol. 35, no. 4, pp. 1469–1484, Oct. 2019, doi: 10.1016/j.ijforecast.2018.10.007.

[35] D. Bahdanau, K. Cho, and Y. Bengio. "Neural machine translation by jointly learning to align and translate." arXiv preprint arXiv:1409.0473 (2014).

[36] T. M. Cover, J. A. Thomas, "Elements of Information Theory, 2nd Edition", Wiley, pp. 80.

[37] "National Centers for Environmental Information," Accessed: Feb. 06, 2023. [Online]. Available: https://www.ncei.noaa.gov/cdo-web/datasets/LCD/stations/WBAN:63819/detail

Yi Hu (Graduate Student Member, IEEE) received the B.S. degree in communication engineering from Chongqing University of Post and Telecommunications, Chongqing, China, in 2014, and the M.S. degree in electrical and communication Engineering from Peking University, Beijing, China, in 2018. Currently he is a Ph.D. candidate in Electrical and Computer Engineering with the Future Renewable Electric Energy Delivery and Management (FREEDM) Systems Center, North Carolina State University, Raleigh, USA. His research interests include smart meter data analysis, and machine learning applications in power distribution systems.

Kai Ye (Graduate Student Member, IEEE) received the B.S. degree in new energy science and engineering from Chinese University of Hong Kong, Shenzhen, China, in 2019, and the M.S. degree in electrical and computer engineering from University of Minnesota, Minneapolis, MN, USA in 2020. Currently he is a Ph.D. candidate in electrical and computer engineering with the Future Renewable Electric Energy Delivery and Management (FREEDM) Systems Center, North Carolina State University, Raleigh, NC, USA. His research

interests include integration of distributed energy resources and data-driven load modeling in distribution systems.

Hyeonjin Kim (Graduate Student Member, IEEE) He received the B.S. degree in electrical engineering from Konkuk University, Seoul, Korea in 2019. He is currently pursuing the Ph.D. degree in electrical and computer engineering with North Carolina State University, Raleigh, NC, USA. His research interests include smart meter data analysis, integration of distributed energy resources, and data-driven model development in distribution systems.

Ning Lu (Fellow, IEEE) received the B.S. degree in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 1993, and the M.S. and Ph.D. degrees in electric power engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 1999 and 2002, respectively. She is a Professor with Electrical and Computer Engineering Department, North Carolina State University. Her research interests include modeling and control of distributed energy resources, microgrid energy management, and applying machine learning methods in energy forecasting, modeling, and control.