Deep Factorization Machine Learning for Disaggregation of Transmission Load Profiles with High Penetration of Behind-The-Meter Solar

Zhenyu Zhao, Daniel Moscovitz, Shengyi Wang, Liang Du, *Senior Member, IEEE*, Xiaoyuan Fan

***Abstract*—The ever-growing integration of distributed energy resources (DERs), especially behind-the-meter (BTM) solar generations, poses imperative operational challenges to system operators such as regional transmission organizations (RTOs). It is important for RTOs to effectively and accurately extract actual load profiles at the transmission level for a single node with significant BTM solar injection. This paper first illustrates the necessity of disaggregating the daily actual load profile of a single node. Furthermore, by segmenting nodes with selected time-series features, nodes with significant BTM solar generation are identified. Lastly, a bi-level framework is proposed, comprising reference node disaggregation and DeepFM nodal disaggregation, aimed at disaggregating the nodal load profiles from which system operators require more information. By adopting a hybrid Deep Factorization Machine (DeepFM) model, the model achieve accurate results by extracting both linear and nonlinear relations between nodes in the same region and the zonal load and nodal load profile. To overcome the lack of ground truth, this paper segments the load profile into daytime, nighttime, and zero-crossing points and utilizes the latter two for evaluation purposes. The proposed disaggregation procedure is validated using real-world, minute-level, normalized, and anonymized nodal data in the PJM service territory.**

***Index Terms*—Factorization Machine, Behind-the-meter Solar, Load Disaggregation**

NOMENCLATURE

Indices and Sets

i	Index of transmission zones
Z_i	Transmission zonal load profile of each zone i
$\mathbf{Z}$	Set of all zonal load profiles
S_i	Proxy solar profile of each zone i
S_i^{MA}	Proxy solar profile with MA
$\mathbf{S}$	Set of all proxy solar profiles
n	Index of transmission nodes in each zone
X_{node_n}	Nodal load profiles of each zone i
$X_{\text{node}_n}^{\text{MA}}$	Nodal load profiles with moving average
$X_{\text{node}_{\text{ref}}}$	Metered nodal load profile of reference node
$P_{\text{node}_{\text{ref}}}$	Net nodal load profile of reference node
$\mathbf{X}$	Set of transmission nodal load profiles
P_{node_n}	Net load profiles of each zone i
$\mathbf{P}$	Set of net transmission nodal load profiles
$\mathbf{X}_{\text{BTM}}$	Set of transmission nodes with BTM solar
$\mathbf{X}_{\text{MA}}$	Set of transmission nodes with moving average
$s_{\text{node}_{\text{ref}}}$	BTM solar of reference node

Parameters and Variables

T_{daytime}	Time period from sunrise to sunset
$T_{\text{nighttime}}$	Time period from sunset to sunrise
T_{alltime}	Whole dataset without segmentation
σ	Standard deviation
C_{l}	Coefficient of net load in linear model
p_{c}	Constant of net load in linear model
C_{ref}	Coefficient of net reference nodal load
p_{ref}	Constant of net reference nodal load
λ	Moving average step size
w_i	Weight term in FM model
w_0	Intercept term in FM model
w_{ij}	weight term in FM model
$\langle \mathbf{v}_i, \mathbf{v}_j \rangle$	Latent factor for features, FM model
$\hat{y}(\mathbf{x})$	Prediction of net nodal load
d	Number of DNN layers
$W^{(d)}$	Weight of DNN model, layer d
$b^{(d)}$	Biases of DNN model, layer d

Functions

$\text{R}(\mathbf{Z}, X)$	Relation between zonal and nodal load profiles
C_{Zn}	Correlation between zonal and node n's load
$f^{(d)}$	Outcome of embedding layer in FM
$D_{\text{W}}(\cdot,\cdot)$	Wasserstein distance
$D_{\text{KL}}(\cdot,\cdot)$	Kullback-Leibler divergence
$D_{\text{KLsym}}(\cdot,\cdot)$	Symmetrical Kullback-Leibler divergence

Manuscript received May 3, 2024; revised September 2, 2024; accepted on November 27, 2024. Date of publication XXX, 2024; date of current version XXX, 2024. L. Du was supported in part by the National Science Foundation (NSF) under Award 2238414. Paper no. 2024-CDPA-0516. (Corresponding author: L. Du.)

A conference version of this paper was presented at ECCE 2023 [1].

Z. Zhao, S. Wang, and L. Du are with the Department Electrical and Computer Engineering, Temple University, Philadelphia, PA 19122, USA. E-mail: {z.zhao, shengyi.wang,ldu}@temple.edu.

D. Moscovitz is with PJM Interconnection, Audubon, PA 19403, USA. E-mail: daniel.moscovitz@pjm.com

X. Fan is with the Pacific Northwest National Laboratory, Richland, WA 99354 USA. E-mail: xiaoyuan.fan@pnnl.gov.

I. INTRODUCTION

Globally, electric power grids are experiencing major paradigm shifts to operate with dominating renewable energy resources. The ever-growing penetration of distributed energy resources (DERs), especially behind-the-meter(BTM) solar generations, has imposed significant impacts on nodal load profiles and consequently poses imperative operational challenges to system operators such as regional transmission organizations (RTOs). BTM resources represent small-scale installations physically located "behind" utility meters or substation telemetry, i.e., without individual sub-metering. California Energy Commission estimated over 40,000 GWh annual BTM solar by 2030, and its impact on RTOs has been overlooked [2].

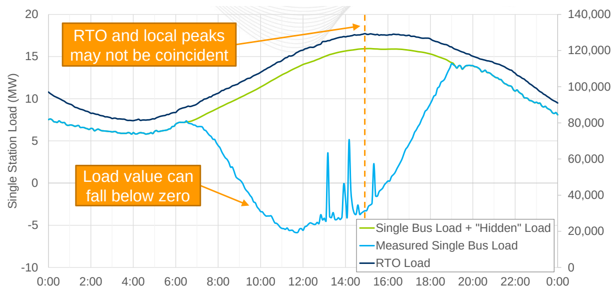


Fig. 1: Real-world example at PJM that needs actual load profiles: the peak load time of the entire RTO (the top line) could be one hour after the peak load time of a transmission (the middle line), whose actual load curve is unknown due to BTM solar (the bottom line) and self-managed loads [3].

Fig. 1 illustrates a real-world example of how BTM solar penetration could impact nodal load profiles and why advanced techniques to extract actual load profiles (i.e., metered or state-estimated load profiles without injection of any BTM resources) are of imminent needs, which represents an actual problem at PJM Interconnection, a major RTO in the U.S and one of the largest transmission system operators in the world. As shown in Fig. 1, an individual node's load profile follows the pattern of RTO load profile during the nighttime. During the daytime, the load profile of single node records a negative power flow during zonal peak load times and remains negative for an extended period. Moreover, with the unknown actual net load curve, the peak load time of the RTO load could be hours later than the single transmission node's peak time, which is expected to be synchronized. As the amount of BTM solar injection increases, causing the duck curve to deepen, the need for ramping-up power also substantially increases, with an unknown magnitude. The actual load remains unknown may also lead to potential over generation.

To maintain the reliability of power systems with high penetration of BTM solar and enhance situational awareness, system operators must be aware of the actual load at each transmission node. Conventionally, utilities conduct random sampling to estimate BTM solar generation in unregistered areas. Various data-driven and machine learning techniques have been proposed recently to dissaggregate BTM solar and actual load at distribution feeders or households. However, effectively extract actual load profiles at transmission nodes with significant BTM solar generation remains a gap in the existing technology paradigm and is of imperative needs for RTOs [4]. Due to the large number of transmission nodes, the size of real-world data, and the need of real-time situational awareness, advanced data-driven approaches are recognized as the enabling scheme with benefits such as eliminating the need for human intervention, reducing costs for estimating actual load profiles, and enabling regular updating of datasets to keep the disaggregation procedure up-to-date.

In general, the energy disaggregation problem for distribution-level households is commonly formulated as reconstructing the aggregated waveform using consumption patterns and model parameters optimized during the learning procedure [5]. To summarize the existing literature, a model-based disaggregation is proposed in [6] for BTM solar and battery storage simultaneously with customized features. On the contract to model-based frameworks, a data-driven-only framework is proposed in [7], presenting a two-layer disaggregation approach for residential levels using only smart meter data, which is similar to the two-layer architecture proposed in [8]. Furthermore, another two-stage decoupled estimation approach is introduced in [9], where the first stage of the model learns the pattern of observable solar power and actual load, and the second stage estimates the solar power and customer load during the demand response event, aiming for accurate aggregated baseline load estimation. Similarly, a model-free energy disaggregation model at the substation level is discussed in [10], utilizing partially labeled data as part of the offline known disaggregation ground truth, which is similar to the semi-supervised setting proposed in this work. Regarding data availability issues, an adaptive framework utilizing partially labeled data with minimal number of BTM PV generation measurement sensors is also constructed in [11]. Similarly, [12] proposes a consumer mixture model to disaggregate BTM solar using consumption patterns of neighboring customers with on solar installations, which is validated with data comes from different aggregation levels. To improve the flexibility of power systems, the authors in [13] propose a machine learning approach to estimate solar capacity accurately from net load data for demand response baseline estimation. Finally, to address the data privacy while using customer's real load data, [14] proposes a federated learning method to disaggregate BTM solar at the community level while preserving the privacy of customers.

However, in the existing literature, most if not all data-driven BTM solar disaggregation are focused on distribution level, and at least have partial labeled data exists as the 'ground truth" for supervised training. In [15], an unsupervised house-level BTM solar disaggregation method is proposed based on physical models and validated with semi-synthetic data. Since it is validated that the relations between zonal and nodal load profiles established during nighttime also hold during daytime [4], as the first effort to disaggregate BTM solar at the transmission level without ground truth, the authors proposed a linear regression-based zonal-to-nodal (Z2N) model utilizing nighttime zonal load and nodal load profiles in [16]. To enhance the performance of Z2N models and alleviate the limitation of model-based approaches in interpretability and accuracy, this paper proposes a data-driven-only deep learning disaggregation framework to comprehensively capture linear and nonlinear relationships among various data sources.

To fully utilize widely available data and capture nonlinear patterns in power systems applications, researchers have proposed a variety of advanced deep learning based comprehensive frameworks to address various power system issues. To localize real-time faults in distribution networks, a deep convolutional neural network classifier is applied to feature vectors extracted from distribution lines [17]. Moreover, a Graph Neural Network (GNN) is adopted to enhance situational awareness and fault location accuracy [18], which can capture the spatial-temporal relationship between data from different sensors in various locations. In [19], [20], several multi-head `Transformer`-based deep learning architectures are adopted for better performance for more accurate solar forecasting to enhance performance in solar generation forecasting.

In this work, to address the limitation of the linear-regression-based approach on nodal load profiles with negative injection and enhance the performance for all single nodes in the same region, a Factorization Machine (FM) learning model [21] is adopted to disaggregate nodal profile utilizing zonal load profile and reference node load profile [1]. The proposed Deep factorization Machine based load disaggregation framework **DFMLD** consists of three modules: data processing and analysis, deep learning, and quantitative evaluation, which is shown in Fig. 2.

- Data processing and analysis: Real-world RTO nodal data is first cleaned by removing days with incorrect readings, erroneous recordings, or interrupted operations, which are usually referred to as "cut-in" and may result in consecutive zeros within the dataset. Next, to further analyze characteristics of each nodal profile, the cleaned data is segmented into daytime and nighttime periods. Each segment, as well as the combined dataset, is used to analyze the correlations between each nodal load profile and the zonal load using the Pearson Correlation Coefficient (PCC) [22]. Standard Deviation is utilized for identifying nodes with high DC-offset.
- Deep learning: To comprehensively capture both linear and nonlinear patterns between these load profiles, a Deep Factorization Machine (DeepFM) is adopted [23]. As a hybrid model, in addition to the linear and bilinear terms in the FM, a deep feed-forward learning component is included to extract nonlinear relationships.
- Quantitative evaluation: Symmetrical Kullback-Leibler (KL) divergence and Wasserstein distance are adopted for evaluating similarity between two time series [24] in a sem-supervised manner, i.e., between disaggregated and actual nighttime profiles (since there is no BTM solar generation at nighttime only).

The technical contributions made by this paper can be summarized as follows.

- A novel Deep-Factorization Machine based load disaggregation framework **DFMLD** is proposed which contains multiple module including data process and quantitative evaluation procedure without known ground truth.
- This paper proposes a comprehensive data cleaning and processing module with defined rules to improve the quality of real-world transmission-level data for deep learning.
- The transmission nodes within the same zone are further segmented into different categories based on profile-related parameters and characteristics. Three different datasets, each containing varying lengths of data which are: 1) daytime data only, 2) nighttime data only, and 3) both daytime and nighttime data—are used within the analysis module, respectively.
- The proposed DFMLD is designed based on observations from real-world RTO data and has been validated by real-world RTO data, which has driven steps forwards industry deployment and adoption into the RTO energy management systems (EMS).

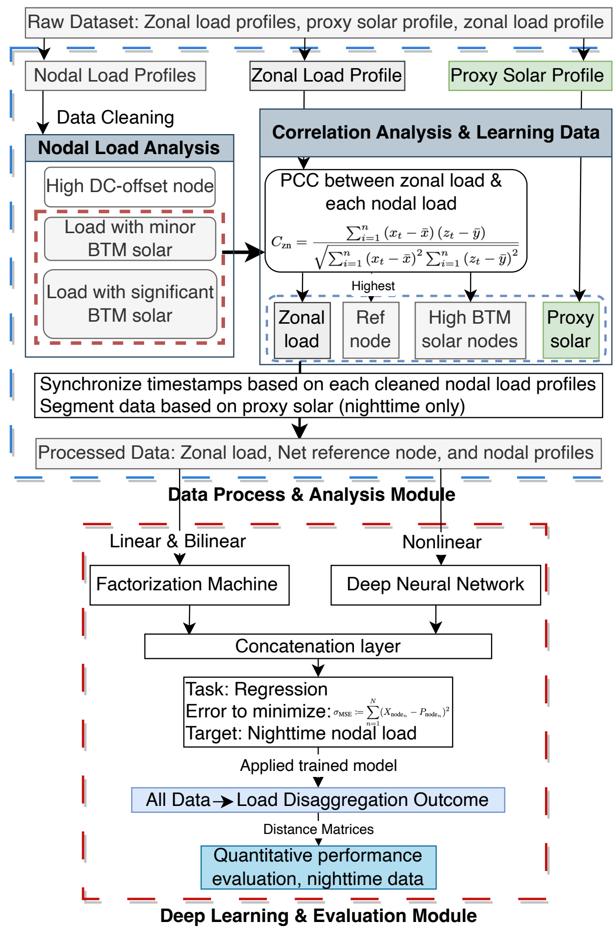


Fig. 2: Pipeline of the proposed **DFMLD** framework.

The remainder of this paper is organized as follows. Section II describes the data formats and resources and then formally defines the energy disaggregation problem considered in this paper. Section III introduces the data processing and analysis module, which will be utilized by the proposed **DFMLD** framework, which is proposed in Section IV in details. Numerical validation results using real-world RTO data are reported in Section V. Finally, Section VI summarizes findings of this work and discusses some potential future directions.

II. Data Description & Problem Formulation

As the framework is formulated in a model-free, data-driven manner, the input data used in calculations and modeling determines the quantitative results. Therefore, it is important to first introduce the data, followed by the problem formulation.

A. *Data Description*

The data utilized in this paper are collected from three types of sources:

- Metered zonal load profiles $\mathbf{Z} = \{Z_1, Z_2, \ldots, Z_i\}$,
- Proxy solar profiles metered in solar farm in corresponding zone $\mathbf{S} = \{S_1, S_2, \ldots, S_i\}$,
- Metered multiple nodal load profiles within each zone $\mathbf{X} = \{X_{\text{node}_1}, X_{\text{node}_2}, \ldots, X_{\text{node}_n}\}$,

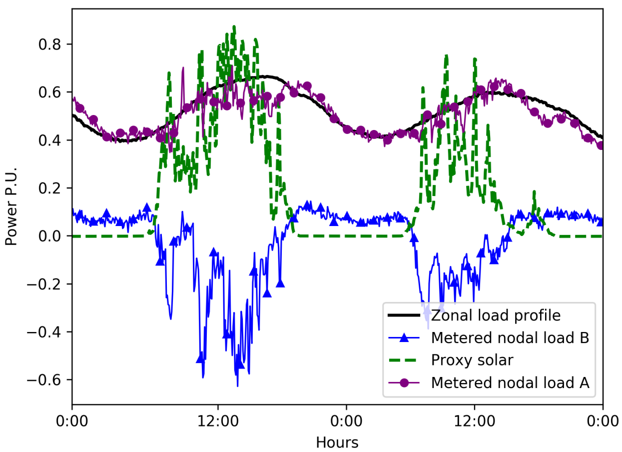


Fig. 3: Illustration of different sources of data real-world metered data. Dotted green line is the proxy solar profile, solid black is the zonal load profile, solid purple line with round markers and solid blue line with triangle markers represent different nodal load profiles.

As illustrated in Fig. 3, which contains 24 hours of data from various sources, including one proxy solar data, one zonal load profile, and two nodal load profiles within the same zone. All of these files from various sources consist of minute-level time-series data that has been anonymized and normalized for confidentiality. It can be observed that

- During nighttime, different nodal load follows the trend of the nodal load profile despite the different in scales.
- During the daytime, with significant BTM solar, both node 1 and node 2 present very different load profiles compared to the zonal total demand.

B. Problem Formulation

The objective of **DFMLD** is to disaggregate actual nodal load profiles $\mathbf{P} = \{P_{\text{node}_1}, P_{\text{node}_2}, \ldots, P_{\text{node}_n}\}$ from metered nodal profiles $\mathbf{X}$. To create a universal procedure for disaggregating all nodes in one zone, the problem is formulated as a multi-module deep learning-based framework. First, the raw data is filtered using pre-defined rules to clean bad data, followed by a moving average (MA) (with step size λ) for data smoothing. The smoothed load and solar profiles are thus

$$\begin{aligned} X^{\text{MA}}_{\text{node}_n}(t) &= \frac{1}{\lambda}(X_{\text{node}_n}(t) + \ldots + X_{\text{node}_n}(t+\lambda-1)), \\ S_{\text{MA}}(t) &= \frac{1}{\lambda}(S(t) + \ldots + S(t+\lambda-1)). \end{aligned} \tag{1}$$

where the moving averaged load profile set is denoted as $\mathbf{X}_{\text{MA}}$. The moving average processing improves not only the computational efficiency but also visualization of time-series data.

The processed data is then segmented into two parts based on the value of proxy solar $\mathbf{S}$: T_{daytime} (with BTM solar), $T_{\text{nighttime}}$ (no solar), and the tuning points (sunrise and sunset). The nodes in the nodal load set $\mathbf{X}$ are further segmented into different groups based on their load patterns, using daytime data only and all-time data, respectively. Also, reference node $X_{\text{node}_{\text{ref}}}$ is selected inside set $\mathbf{X}$ which is the nodal load has highest correlation with zonal load $\mathbf{Z}$ and net load of reference node $P_{\text{node}_{\text{ref}}}$ is acquired through existing Z2N procedure by

$$p(t) = C_l Z(t) + p_c + \epsilon(t), \quad \forall t = 1, \ldots, T, \tag{2}$$

where the actual nodal load profile $\mathbf{p}$ follows the zonal load profile data $\mathbf{z}$ in an affine manner, subject to a constant weight C_l, a constant load component p_c, and an unknown stochastic load mismatch error ϵ [16]. As a regression problem, parameters are trained by minimizing the difference between metered and disaggregated nodal load profiles:

$$\arg\min \| P_{\text{nodal,nighttime}}, \hat{P}_{\text{nodal,nighttime}} \|. \tag{3}$$

The primary challenge is that there is no ground-truth for disaggregation outcomes of $\mathbf{P}$. To accurately disaggregate the actual nodal load and overcome the lack of ground-truth, the disaggregation procedure is based on the relation between transmission level zonal and nodal load which is the nighttime relations still stands during daytime

$$\text{R}_{\text{daytime}}(\mathbf{Z}, X) = \text{R}_{\text{nighttime}}(\mathbf{Z}, X) \pm \varepsilon, \quad \forall X \in \mathbf{X} \tag{4}$$

where ε represents an unknown transposition error term which combines stochastic mismatch errors from both the load and BTM solar and needs to be learned as a residue term [4]. To obtain the trained relations necessary for disaggregating the net nodal load, this paper proposes a novel learning and evaluation procedure based on data segmentation and a hybrid model capable of capturing linear, bilinear, and nonlinear relations.

Specifically, each zonal load profile $\mathbf{Z}$ and corresponding nodal load profiles in set $\mathbf{X}$ are partitioned into three segments based on proxy solar :

- When $S < 0$, the segmented data represents nighttime;
- When $S > 0$, the segmented data represents daytime;
- When $S = 0$, the segemtned data represents a set of cross-zero points (sunset and sunrise times).

The nighttime data of zonal load $\mathbf{Z}$ and net reference nodal load $P_{\text{node}_{\text{ref}}}$ are utilized for training to disaggregate all nodal load profiles in set $\mathbf{X}$:

$$P_{\text{node}_n} = f(P_{\text{node}_{\textbf{ref,nighttime}}}, \mathbf{Z}_{\text{nighttime}}), \tag{5}$$

where f is learned from the the proposed deep learning module. Note that the nighttime data is utilized not only during the training but also for evaluating the disaggregation outcomes of $\mathbf{P}$ since there is no BTM solar generation during the nighttime, which results in $X_{\text{node}_n} = P_{\text{node}_n}$. In other words, the nighttime data can therefore be used for disaggregation evaluation as well [4]. Moreover, the learned function f is then applied to the entire dataset including both daytime and nighttime segments, as the relations established during the night also hold during the daytime [4]. The boundaries between daytime and nighttime data can be represented by the sunrise and sunset events, which can in turn be considered as cross-zero points, i.e., at the intersection of daytime and nighttime segments and thus belong to both. Therefore, these points are utilized in both the Z2N procedure for performance enhancement and the evaluation step to overcome the challenge of no (or only partial) ground truth.

III. Data Processing and Analysis Module

Within the data processing and analysis module, pre-defined rules are applied for data cleaning. The nodal load profiles are then further segmented, and a reference node is selected.

A. Data Cleaning Module

Lack of data quality in the above domains manifests in several forms, including missing, incomplete, inconsistent, inaccurate, duplicate, and outdated data, which can result in low performance when utilized in machine learning and analytical models. While using real-world transmission network nodal load profiles and proxy solar data in the proposed data-driven techniques, uncertainty and error in measurement and communication could causes random noises [25]. The following abnormalities are observed in the real-world dataset:

- The load profile sudden drops to zero during the night (when no BTM solar injection), followed by recoveries within seconds or sometimes minutes to hours.
- The load profile suddenly drops from non-zero to zero with an abnormal trend during daytime.
- The proxy solar suddenly drops to zero during daytime while the positive readings are expected.
- Null values (NaN) have been observed in both proxy solar and load profiles, range from one minutes to hours.

These data abnormalities may be caused by different reasons, including but not limit to data transmission error, relay devices malfunction, as well as services on a branch (line, transformer, series device, or phase shifter) which could affect modeled topology, device reconfiguration, or equipment upgrade which causes no readings for an extended period. Therefore, it is of great importance to clean the data to improve its quality first. There are two sets of data that need to be looked into: the load profile dataset $\mathbf{X}$ and the proxy solar dataset $\mathbf{S}$, as no data abnormality has been observed in zonal load profiles. As illustrated in Section II, all data is processed on a daily basis, and so is the cleaning procedure. For daily load and solar files segmented into daytime and nighttime data, clean the data based on two novel pre-defined rules. Rule 1 cleans the data based on abnormal detection during nighttime. It removes the entire day's data if any zero value exists in the nighttime data. This indicates a recording error since nighttime load should be greater than zero, as there's no BTM solar generation. While rule 2 cleans the data using daytime data, the rule defines that if there are consecutive 14 data points (20 minutes) of zero, the daily profile should be deleted. This indicates either a cut-in procedure or a reading error, which will result in invalid outcomes in the nodal analysis module. The threshold of 14 is chosen based on the deep learning model's robustness against missing data while the missing data is less than 1% [26]. The two rules are listed below:

$$\text{If} \ \forall \ X_{\text{nighttime}}[t] = 0, \ \text{delete the day,} \qquad \text{(Rule 1)}$$

$$\text{If} \sum_{t=i}^{i+14} |X_{\text{daytime}}[t]| = 0 \ \text{for any} \ i, \text{delete the day.} \qquad \text{(Rule 2)}$$

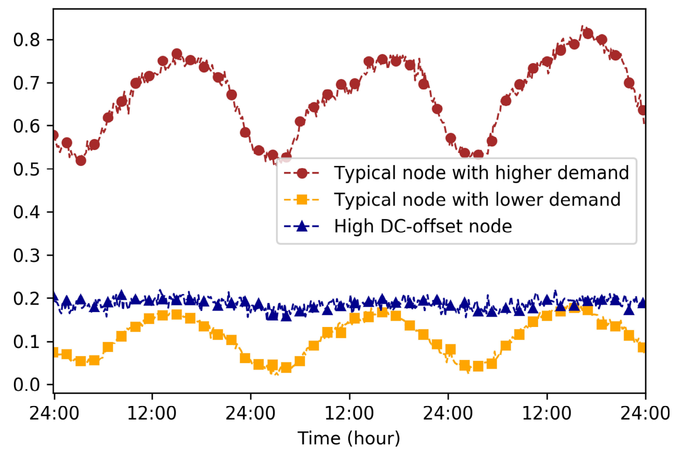


Fig. 4: Illustration of different nodes with significantly different patterns, where the dark blue line with triangle markers represents the pattern of a node with high percentage of DC-offset.

B. Nodal Load Profile Analysis

In each zone, $\mathbf{X}$ contains distinct groups of nodal load profile patterns. Before disaggregating nodal load profiles, all nodes need to be analyzed to identify which nodal load profiles require disaggregation. With the increasing proportion of renewable energy generation, load profiles will significantly change if solar is behind-the-meter. During the daytime when solar generation peaks, the nodal peak time will be delayed for hours, and the trend will show a significant difference, which is expected to be similar. As shown in Fig. 3, the purple line with round markers exhibits subtle value drop observed during the solar generation peaks, while the blue line with triangle markers shows a significant drop and even goes under zero during noon.

Besides the aforementioned observations that nodes could present very different patterns at varying levels of BTM solar penetration, nodes with a substantial percentage of DC offset have a unique profile pattern. As shown in Fig. 4, it can be observed that different from the typical load, the orange dotted line with square markers and the brown dotted line with the round markers, the dotted dark blue line with triangle markers presents no peaks or troughs as the value remains relatively flat and only fluctuates within a small range.This is attributed to a high percentage of load consumed by always-on facilities, such as data centers.

Therefore, there are three types of nodal loads: nodes without BTM solar injection, nodes with BTM solar injection, and nodes with high DC-offset. Therefore, the nodal load profile analysis module serves three purposes:

- Identify nodes with or without significant BTM solar (both daytime and nighttime data utilized).
- Identify high DC-offset nodes (all-time data included).
- Evaluate the similarity between nodal load profiles and corresponding zonal load profiles, and select a reference node to be further used in the disaggregation module (nighttime data utilized).

To maximize the utilization of the dataset, various segments of the load profile are utilized for analysis purposes, including daytime-only, nighttime-only, and all-time data. To identify high DC-offset nodes, *Standard Deviation* is adopted to evaluate the fluctuation of time-series data, formulated as:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_t - \mu)^2}{N}}, \tag{6}$$

where x_t represents each individual data point in each dataset, μ is the mean (average) of each dataset, and N is the total number of data points. If the value of single file is relatively low compared to overall of other files inside a dataset, it indicates that the data fluctuates within a small range. All-time data is used for evaluation. To analyze the relations between the zonal load **Z** and nodal load X_i without negative injection, the Pearson Correlation Coefficient (PCC) is used to quantitatively evaluate the correlation using nighttime data:

$$C_{\text{zn}} = \frac{\sum_{i=1}^{n} (x_t - \bar{x})(z_t - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_t - \bar{x})^2 \sum_{i=1}^{n} (z_t - \bar{y})^2}}, \tag{7}$$

which x_t and z_t represent each nodal load profile and zonal load profile, respectively. n represents the number of data points. For which a larger absolute value of C_{zn} indicates stronger linear relations. The nodal load profile with the highest value of PCC is selected as the reference node for the corresponding zonal load.

To identify how many files in **X** have substantial BTM solar injection, utilize the observed patterns of the impact of solar penetration. These patterns include significant declines during the daytime and increased ramp-up energy required during sunset. To classify nodes, PCCs are used to evaluate the relation changes during the daytime and nighttime. For nodes without BTM solar injection, the relations between the nodal load profile and the corresponding zonal load profile during daytime and nighttime are in an affine manner [4]:

$$\frac{\text{R}_{\text{daytime}}(\text{Z}, X)}{\text{R}_{\text{nighttime}}(\text{Z}, X)} = (1 \pm 0.05), \tag{8}$$

where the relations during daytime and nighttime are bounded within the range of $\pm 5\%$ error, which is widely acceptable in RTO operations. Therefore, for a single node in **X** with a day-night ratio greater than 1.1, the nodes are classified as having BTM solar injection and included in dataset $\mathbf{X}_{\text{BTM}}$.

IV. DeepFM Based Learning Module

In the DeepFM-based learning module, net load profile of the reference node is first acquired by the Z2N disaggregation process [16], which utilizes the nighttime zonal load denoted as Z, and the nighttime metered reference nodal load profile denoted as X_{noderef}, to disaggregate the actual reference load denoted as $P_{\text{node}_{\text{ref}}}$, by formulating it as a regression task to acquire the relation. After acquiring the disaggregated reference load $P_{\text{node}_{\text{ref}}}$, along with the zonal load Z, the variables are used as the input for the hybrid deep learning model to disaggregate the actual node in $\mathbf{X}_{\text{BTM}}$. To acquire linear, bilinear, and nonlinear relationships, all parameters are obtained through training on the nighttime data.

Algorithm 1: Deep Factorization Machine Learning Based Load Disaggregation Framework

Input S, Z, λ, **X**
`// Applying Moving Average`
$X^{\text{MA}}_{\text{node}_n} \leftarrow \frac{1}{\lambda}\big(X_{\text{node}_n}(t) + \ldots + X_{\text{node}_n}(t+\lambda-1)\big)$
$S_{\text{MA}}(t) \leftarrow \frac{1}{\lambda}(S(t) + \ldots + S(t+\lambda-1))$
for $t = 0 : T-1$ **do**
  $X^{\text{MA}}_{\text{node}_n}(t+1) \leftarrow X^{\text{MA}}_{\text{node}_n}(t) - X_{\text{node}_n}(t) + X_{\text{node}_n}(t+\lambda)$
  $S_{\text{MA}}(t+1) \leftarrow S_{\text{MA}}(t) - S(t) + S(t+\lambda)$
  $t \leftarrow t+1$
end
`// Nodal Load Profiles Analysis`
Input Z, $\mathbf{X}_{\text{MA}}$, Initialize variables
for *each X in dataset* $\mathbf{X}_{MA}$ **do**
  Calculate $T_{nighttime}$ zone-node correlation
   Calculate day-night ratio
  if $X_{node_{ref}}$ *zone-node PCC is the highest* **then**
    Select the highest as reference node
  end
  if $\forall$ X *day-night ratio* > 1.1 **then**
    Add to set $\mathbf{X}_{\text{BTM}}$
  end
end
Output $X_{\text{node}_{\text{ref}}}$, $\mathbf{X}_{\text{BTM}}$
`// Disaggregation at Reference Node`
Input $\text{X}_{\text{node}_{\text{ref}}}$ $(T_{nighttime})$, Z
Calculate $C_{\text{ref}}, p_{\text{ref}} \leftarrow$ Min. σ_{MSE}, OLS outcome
Update $C_{\text{ref}}, p_{\text{ref}} \leftarrow$ Min. switching points error, tuning
Output $P_{\text{node}_{\text{ref}}} = C_{\text{ref}} Z + p_{\text{ref}}$
`// Deep FM Disaggregation`
Input response variables $X_{\text{node}_{\text{ref}}}$, Z, target variable $\mathbf{X}_{\text{BTM}}$
Training task: regression
for $n = 1, \ldots, N, n \neq$ *ref* **do**
  Features: $X_{\text{node}_{\text{ref}}}$ $(T_{nighttime})$, Z $(T_{nighttime})$
  Regression target: X_{node_n} $(T_{nighttime})$
  Train Hyper-parameters $\leftarrow$ Min. nighttime σ_{MSE}
end
Output trained model
Apply trained model on data $(T_{alltime})$

A. Reference Node Selection and Disaggregation

The first step of the proposed framework is to identify the corresponding reference node. Specifically, the reference node of each dataset is chosen based on the Pearson Correlation Coefficient (PCC) between the nodal and zonal load profiles at nighttime. The node has the highest PCC value is identified as the reference node node_{ref}. As the actual load profile of reference node is required in the deep learning module, the reference node is then disaggregated to the actual load profile of reference node $P_{\text{node}_{\text{ref}}}$ by applying the Z2N disaggregation process, as justified in [4]. On the reference node, where both $P_{\text{node}_{\text{ref}}}(t)$ and $s_{\text{node}_{\text{ref}}}(t)$ are unknown. The disaggregation for the reference node is formulated based on Eqn. (2):

$$P_{\text{node}_{\text{ref}}}(t) = C_{\text{ref}} Z(t) + p_{\text{ref}} + \epsilon(t), \quad \forall t = 1, \ldots, T, \tag{9}$$

where ϵ represents an unknown stochastic load mismatch error. The constant weight C_{ref} and constant load component p_{ref} are

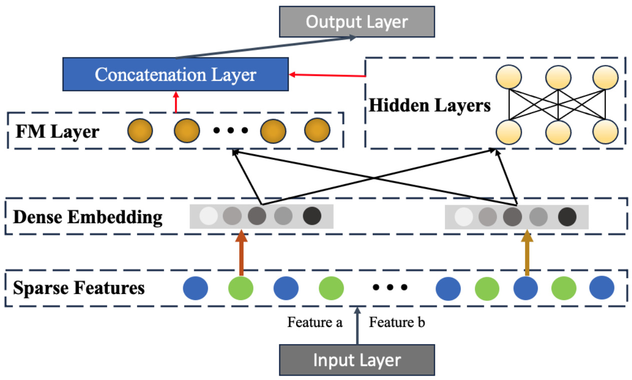


Fig. 5: Detail structure of Deep Factorization Machine learning module

obtained by by minimizing the error between the actual and disaggregated nodal load profiles using Ordinary Least Squares (OLS) [27]:

$$\arg\min \frac{1}{t}\sum_{i=1}^{t}\left(\mathbf{P}_{\text{node}_{\text{ref}},\text{nighttime}(t)} - \hat{\mathbf{P}}_{\text{node}_{\text{ref}},\text{nighttime}(t)}\right)^2. \quad (10)$$

The nighttime data is used for training purpose. Noted that during nighttime, the actual reference node load $P_{\text{ref},\text{nighttime}(t)}$ equals to the metered reference node load $X_{\text{ref},\text{nighttime}(t)}$. To enhance the obtained disaggregation performance, the accumulated errors on both zero-crossing points is minimized:

$$\arg\min \sum_{i=1}^{n}\left(\mathbf{P}_{\text{node}_{\text{ref}}(t)}\right), t \in \{T_{\text{daytime}}, T_{\text{nighttime}}\}, \quad (11)$$

where the final constant weight C_{ref} and constant load component p_{ref} is obtained.

B. DeepFM Learning for Disaggregating All Nodes

The DeepFM module is proposed to to disaggregate all single nodes. The entire module includes multiple layers for feature extraction and learning both high and low orders of relations. Dense embedding layer process the dense input and compress the input vector. This compressed vector is then passed to both the Factorization FM [21] and the Deep Neural Network (DNN), which share the same input. The concatenation layer combines the outputs from each learning module and produces the final output. The details of the hybrid learning model is illustrated in Fig. 5. The key advantages DeepFM module brings are as follow:

- Both low-order feature interactions and high-order feature interactions are considered in FM and DNN, respectively. Linear, bilinear, and nonlinear relations are all captured [23].
- DeepFM training procedure is efficient as different modules receive the same embedded vector.

1) Data Input and Dense Embedding: The response variable input of the model include reference node actual load $P_{\text{node}_{\text{ref}}}$ and zonal load $\mathbf{Z}$, Where the actual reference node load is $P_{\text{node}_{\text{ref}}}$ is acquired by Z2N procedure. The dense inputs of two time-series datasets are compressed into low-dimensional, dense real-value vectors for feature extraction [23]. The same vector is passed to both of the learning modules. The target is individual nodal load profiles in $\mathbf{X}_{\text{BTM}}$.

2) Factorization Machines and Deep Neural Networks: Factorization Machines learning was originally proposed for collaborative recommendation [21], and it provides substantial enhancements over linear regression as well [28]. Within FM, the model includes an intercept term denoted as w_0 and a weight parameter w_i associated with each feature x_i which has corresponding n numbers of weights. FM estimates the target variable by capturing the interactions between all pairs of features using factorized interaction parameters [28]. The predicted value for the target variable $\hat{y}(\mathbf{x})$is calculated by

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n} w_{ij} x_i x_j. \quad (12)$$

In this paper, the feature dimension i is 2, which is the zonal load profile $\mathbf{Z}$ and the actual load profile of reference bus $P_{\text{node}_{\text{ref}}}$. To address the issue of computational complexity, latent factors $\langle \mathbf{v}_i, \mathbf{v}_j \rangle$ associated with each feature are introduced [21]. The mathematical representation of the FM equation becomes

$$\hat{y}(\mathbf{x}) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle x_i x_j. \quad (13)$$

Both linear and bilinear terms are utilized to capture the relationship between features and the target. The deep learning module is a feed-forward deep neural network consisting of all the training components with FM. Denote the outcome of the embedding layer:

$$f^{(0)} = [T_1, T_2], \quad (14)$$

where T_i include $\mathbf{Z}(t)$ and $P_{\text{node}_{\text{ref}}}(t)$. Then $f^{(0)}$ is fed into the deep neural network, and the forward process is:

$$f^{(d+1)} = \sigma\left(W^{(d)} f^{(d)} + b^{(d)}\right), \quad (15)$$

where d represents the layer depth, which is 4 in this paper, and σ denotes the activation function, where ReLU is adopted in this model. $f^{(d)}$, $W^{(d)}$, and $b^{(d)}$ respectively denote the output, model weights, and biases of the d-th layer. After that, a dense real-value feature vector is generated. The parallel structure of leaning components brings several benefits:

- FM model serves as part of the overall learning architecture, there's no need to use FM's latent feature vectors to initialize the deep neural networks.
- The low and high order features interactions are learned from raw data to reduce information loss.
- No separate feature engineering is needed for the deep learning component [23].

3) Concatenation and Output Layer: For the target node in the same region node$_n$, the outputs of both learning components are concatenated in a concatenation layer, resulting in matrix C with dimensions $(n, k + d)$, where n is the number of samples, k is the number of latent factors from the FM component, and d is the dimensionality of the output of the DNN component. The hyper-parameters are trained by minimizing the Mean Square Error (MSE) between the disaggregated results P_{node_n} and metered load X_{node_n} with a supervised setting, which is denoted as:

$$\min \quad \sigma_{\text{MSE}} \coloneqq \sum_{n=1}^{N}(X_{\text{node}_n} - P_{\text{node}_n})^2. \tag{16}$$

The comparison of the proposed framework by this paper with the state-of-the-art is summarized in the Table I.

V. Numerical Validations

The proposed framework is validated using real-world minute-level, anonymized, and normalized PJM data. Two datasets of differene sizes are utilized to validate the performance of **DFMLD** in different situations.

- Dataset A contains a set of 2 transmission nodal load profiles, 1 zonal load profile, and 1 proxy solar profile. Which is selected by PJM with one nodal load with negative injection during nighttime and the other node has substantial BTM solar yet remains positive value during daytime;
- Dataset B contains a set of 40 transmission nodal load profiles (after cleaning), 1 zonal load profile, and 1 proxy solar profile.

Each file is year-long, normalized, and anonymized. Dataset A is pre-processed and cleaned; one of the transmission nodal load profiles is highly correlated to the zonal load, and the other one is a single node with a high percentage of BTM solar penetration, which leads to an extended period of negative metered value during the daytime. Dataset B is subjected to the data cleaning and analysis module.

Remark V.1. Note that real-world load and BTM solar patterns change significantly year-to-year. For example, actual data in 2023 in general look very different from the year of 2022. Therefore, only recent years' data are useful. Moreover, though PJM has 86,397 internal electrical nodes with 10,251 nodes have actual loads, majority of nodes with loads have missing or inconsistent data. We ran various steps and procedures to verify data correctness and completeness and Dataset B is the verified, trustworthy dataset with actual PJM nodes with both loads and significant BTM solar.

A. Data Cleaning and Analysis

The data cleaning procedure is applied to the transmission nodal load profile set and the proxy solar file. After data cleaning and timestamp synchronization, the average data length shrinks from 365 days to 217. To classify the high DC-offset node, `Standard Deviation` is applied, and the distribution is shown in Table II. As shown in the table, a clear threshold of 0.08 separates files into two groups. After manually inspecting the dataset, files with a standard deviation value lower than 0.08 are confirmed as loads with high DC-offset, which matches the quantitative outcomes. Then, the daytime-nighttime zone-node ratio is calculated using the first week of August as the solar strongest period: $\frac{\text{R}_{\text{daytime}}(\text{Z},X)}{\text{R}_{\text{nighttime}}(\text{Z},X)}$.

All 32 nodes have a value less than 0.95, indicating that all nodes have a certain level of solar injection. Moreover, the load profile with highest PCC with zonal alod profile is selected as reference node for deep learning module which is 0.943.

B. Deep Learning Disaggregation Outcomes

The performance of **DFMLD** and the comparison of benchmark disaggregation frameworks are shown in Fig. 6. It can be observed that:

- During nighttime, the **DFMLD** disaggregation procedure follows the metered nodal load more closely than the other two outcomes, which indicates better performance.
- The disaggregation results of **DFMLD** are close to the metered nodal load profile at cross-zero points, corresponding to the sunset and sunrise times, which indicates enhanced performance.
- The disaggregated outcome during daytime follows the trend of the zonal load, consistent with the observations made during nighttime.

Overall, the hybrid model based **DFMLD** outperforms the benchmark Z2N disaggregation procedure and the Factorization Machine based disaggregation procedure.

C. Quantitative Evaluation of Outcomes

To address the challenge of lacking ground truth data, the disaggregated nighttime nodal profile from the proposed **DFMLD** model is utilized to evaluate performance since there is no BTM solar generation during the night. As illustrated in Fig. 6, the performance evaluation focuses on how closely the disaggregated nodal load profile follows the metered nodal load profile. To quantitatively evaluate the how closely it follows, the area-based spatial differences between two time series of equal length is adopted [29]. The similarity between two time-series data with same length is quantitatively evaluated using the accumulation of point-to-point Euclidean distances. Compared to the commonly used MSE, the area difference evaluates the performance by considering the entire time-series data as a whole, instead of only using several timestamps. The smaller the value is, the higher the similarity between two equal length time-series and vice versa. Beyond the area difference, two more distance functions are also utilized for comprehensively evaluating the outcomes.

1) Kullback–Leibler Distance: The Kullback-Leibler (KL) divergence is a statistical measurement of the difference between a discrete (actual) probability distribution P from its reference probability distribution Q in probability theory and information theory, where discrete probability distributions P and Q defined on the same probability space $\mathcal{Y}$. The KL divergence from Q to P is formulated as:

$$D_{\text{KL}}(P\|Q) = \sum_{y\in\mathcal{Y}} P(y)\log\left(\frac{P(y)}{Q(y)}\right), \tag{17}$$

which is the expected logarithmic difference between P and Q. Though KL distance is originally used over two probability distribution, however recent works justified utilizing

TABLE I: Comparison of Different Models (Newly Proposed in This Work vs. State-of-the-art)

Model Name	Learning Algorithm	Model Type	Target	Difference/Notes
Z2N Disaggregation	OLS Regression	Linear only	Single node	First transmission BTM solar disaggregation model
FM-based Disaggregation	Factorization Machine	Linear & Bi-linear	All nodes	Bi-linear model to disaggregate all nodes, same zone
DFMLD	Deep Factorization Machine	Deep Learning	All nodes	DL model for better disaggregation performance

TABLE II: Distribution of Standard Deviation

Standard Deviation Range	Frequency
0 - 0.05	2
0.05 - 0.08	6
0.08 - 0.1	0
>0.1	32

a symmetrical KL divergence to evaluate the time-series' similarity [30], which formulated by

$$D_{\text{KLsym}} = \frac{1}{2}\left[D\left(\mathbf{p}_{\text{nodal,nighttime}} \| \hat{\mathbf{p}}_{\text{nodal,nighttime}}\right) + D\left(\hat{\mathbf{p}}_{\text{nodal,nighttime}} \| \mathbf{p}_{\text{nodal,nighttime}}\right)\right], \quad (18)$$

where $\hat{\mathbf{p}}$ and $\mathbf{p}$ denote the disaggregated and metered nodal load profiles, respectively.

2) Wasserstein Distance: Wasserstein distance metric is originally designed for evaluating between two probability distributions and has been widely used in calculating ambiguity sets in power system applications [31]. The Wasserstein metric $D_W(P,Q): \mathcal{M}(\Xi) \times \mathcal{M}(\Xi) \to \mathbb{R}$ between two distributions P and Q is formulated as:

$$D_W(P,Q) = \inf\left\{\int_{\Xi\times\Xi} \|\boldsymbol{w}_1 - \boldsymbol{w}_2\| \, \Pi\left(d\boldsymbol{w}_1, d\boldsymbol{w}_2\right)\right\}, \quad (19)$$

where $\mathcal{M}(\Xi)$ represents the set of probability distributions with support Ξ, Π is a joint distribution of $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ with marginal distributions P and Q, respectively, and $\|\cdot\|$ is a norm term [4]. In this paper, metered nodal load profile and disaggregated nodal load profile are considered two samples with same length and same support, which matches the best practices in industry.

3) Quantitative Outcomes: The quantitative outcome of the distance matrices is shown in Table III, where the distances are calculated on a three-day basis (4320 data points). The outcomes represent the mean value of all three-day outcomes for each nodal load profile within the corresponding dataset. It can be observed that on both datasets, **DFMLD** performs better than FM (both outperforms Z2N) evaluated by smaller lower area difference, KL distance, and Wassertein distance. Specifically, on Dataset A (nodes with significant BTM solar) **DFMLD** presents significant improvement compared to other two methods, while on Dataset B both FM and **DFMLD** outcomes have less significant improvements and **DFMLD** exhibits slightly better performance compared to the FM model with a narrow margin. The differences in performance improvement could mainly be caused by different scales of BTM solar in each node. Dataset A has significant BTM solar injection (over 100% during some summer days), indicating significant different load patterns of nodal and zonal load profiles during nighttime and daytime. On the other hand, Dataset B has less BTM solar injection (less than 30%) which exhibit higher similarity between zonal and nodal load profiles during daytime. Therefore, the benchmark (Z2N) disaggregation framework has a better outcome with Dataset B as the scale of BTM solar is much smaller, leaving less room for improvement.

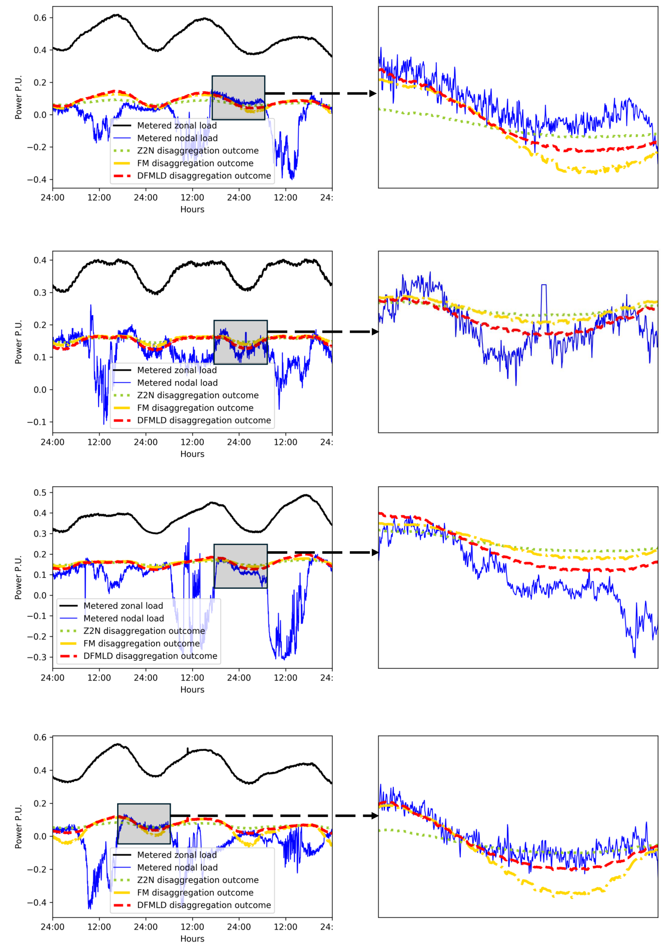


Fig. 6: Comparison of the performance of Z2N (green dotted line), FM disaggregation (yellow dashed line), and **DFMLD** (red dash dotted line) shows significant improvements on small-scale load with negative injection compared to the benchmark during different seasons.

VI. Conclusion

This paper proposed a deep Factorization Machine Learning-based framework (**DFMLD**) to disaggregate actual transmission nodal load profiles from metered profiles with significant BTM solar penetration. Firstly, a comprehensive data clean process was introduced for large batches of real-world data to improve data quality for granular analysis. Secondly, the Pearson Correlation Coefficient served two purposes: 1) the node with the highest coefficient with the zonal load profile during nighttime was selected as the reference, which was then used along with the zonal load for disaggregating all nodes in the same region; 2) daytime-nighttime ratio

TABLE III: Comparison of numerical results on different dataset with different disaggregation models

Dataset	Disagg. Framework	Area Diff.	KL_{SYM} Distance	Wassertein Distance	Area Diff. Improvement	KL_{SYM} Improvement	Wassertein Improvement
Dataset A	Z2N [4]	85.23	52.07	0.031	Benchmark		
	FM	50.56	36.05	0.027	40.6%	30.8%	12.9%
	DFMLD	40.27	27.85	0.025	52.8%	46.5%	19.4%
Dataset B	Z2N [4]	130.62	61.46	0.036	Benchmark		
	FM	81.53	39.45	0.027	37.6%	35.81%	25%
	DFMLD	80.53	39.07	0.027	38.34%	36.43%	25%

between nodal and zonal load was calculated, which could help to distinguish the level of BTM solar penetration. Finally, a hybrid Factorization Machine-Deep Neural Network model was proposed to capture the linear, bilinear, and nonlinear relations between the selected time-series. Validated on real-world data from PJM, this model outperforms both linear disaggregation frameworks and FM-only frameworks.

The proposed DFMLD is designed based on observations from real-world RTO data and has been validated by real-world RTO data. Therefore, the proposed proposed DFMLD is readily available for industry adoption to provide granular modeling of both BTM DERs at scale to enhance system security and reliability under high penetration of DERs.

In terms of limitations, although achieved significant performance improvements, the proposed disaggregation framework has not yet fully utilized the available proxy solar data. Moreover, the performance improvement on nodes with less BTM solar can be further investigated. Suggested future research areas could be focused on not only use the proxy solar as a tool for segmentation and tuning but also consider adopting advanced machine learning algorithms to further enhance the BTM disaggregation performance by mining the relationship between BTM solar and proxy solar.

References

[1] Z. Zhao, D. Moscovitz, L. Du, and X. Fan, "Factorization machine learning for disaggregation of transmission load profiles with high penetration of behind-the-meter solar," in *2023 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2023, pp. 1278–1282.

[2] X. Fan, D. Moscovitz, L. Du, and W. Saad, "A data-driven democratized control architecture for regional transmission operators," in *IEEE PES Innovative Smart Grid Technologies Conf. (ISGT)*, 2022.

[3] D. Moscovitz, "Need for BTM Data: Transmission Outage Analysis," in *The 9th Annual Monitoring and Situational Awareness Technical Conference*. NERC, October 2021. [Online]. Available: https://www.nerc.com/pa/rrm/Resources/Documents/2021_MSA_Conference_Session2.pdf

[4] D. Moscovitz, Z. Zhao, L. Du, and X. Fan, "Semi-supervised, non-intrusive disaggregation of nodal load profiles with significant behind-the-meter solar generation," *IEEE Transactions on Power Systems*, vol. 39, no. 3, pp. 4852–4864, 2024.

[5] M. Khodayar, J. Wang, and Z. Wang, "Energy disaggregation via deep temporal dictionary learning," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 5, pp. 1696–1709, 2019.

[6] F. Wang, X. Ge, Z. Dong, J. Yan, K. Li, F. Xu, X. Lu, H. Shen, and P. Tao, "Joint energy disaggregation of Behind-the-Meter PV and battery storage: A contextually supervised source separation approach," *IEEE Transactions on Industry Applications*, vol. 58, no. 2, pp. 1490–1501, 2022.

[7] F. Bu, R. Cheng, and Z. Wang, "A two-layer approach for estimating behind-the-meter pv generation using smart meter data," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 885–896, 2022.

[8] A. Stratman, T. Hong, M. Yi, and D. Zhao, "Net load forecasting with disaggregated behind-the-meter pv generation," *IEEE Transactions on Industry Applications*, vol. 59, no. 5, pp. 5341–5351, 2023.

[9] K. Li, J. Yan, L. Hu, F. Wang, and N. Zhang, "Two-stage decoupled estimation approach of aggregated baseline load under high penetration of behind-the-meter PV system," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 4876–4885, 2021.

[10] W. Li, M. Yi, M. Wang, Y. Wang, D. Shi, and Z. Wang, "Real-time energy disaggregation at substations with behind-the-meter solar generation," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2023–2034, 2020.

[11] R. Saeedi, S. K. Sadanandan, A. K. Srivastava, K. L. Davies, and A. H. Gebremedhin, "An adaptive machine learning framework for behind-the-meter load/pv disaggregation," *IEEE Trans. Industrial Informatics*, vol. 17, no. 10, pp. 7060–7069, 2021.

[12] C. M. Cheung, S. R. Kuppannagari, A. Srivastava, R. Kannan, and V. K. Prasanna, "Behind-the-meter solar generation disaggregation at varying aggregation levels using consumer mixture models," *IEEE Transactions on Sustainable Computing*, vol. 8, no. 1, pp. 43–55, 2022.

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

[14] J. Lin, J. Ma, and J. Zhu, "A privacy-preserving federated learning method for probabilistic community-level behind-the-meter solar generation disaggregation," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 268–279, 2021.

[15] K. Pan *et al.*, "An unsupervised data-driven approach for behind-the-meter photovoltaic power generation disaggregation," *Applied Energy*, vol. 309, p. 118450, 2022.

[16] Z. Zhao, D. Moscovitz, S. Wang, X. Fan, and L. Du, "Semi-supervised disaggregation of daily load profiles at transmission buses with significant behind-the-meter solar generations," in *ECCE 2022, Detroit, MI*.

[17] M. Zhao and M. Barati, "A real-time fault localization in power distribution grid for wildfire detection through deep convolutional neural networks," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 4316–4326, 2021.

[18] M. MansourLakouraj, H. Hosseinpour, H. Livani, and M. Benidris, "Waveform measurement unit-based fault location in distribution feeders via short-time matrix pencil method and graph neural network," *IEEE Trans. Industry Applications*, vol. 59, no. 2, pp. 2661–2670, 2023.

[19] S. Ziyabari, L. Du, and S. K. Biswas, "Multibranch attentive gated resnet for short-term spatio-temporal solar irradiance forecasting," *IEEE Trans. Industry Applications*, vol. 58, no. 1, pp. 28–38, 2022.

[20] S. Ziyabari, Z. Zhao, L. Du, and S. K. Biswas, "Multi-branch resnet-transformer for short-term spatio-temporal solar irradiance forecasting," *IEEE Transactions on Industry Applications*, vol. 59, no. 5, pp. 5293–5303, 2023.

[21] S. Rendle, "Factorization machines," in *2010 IEEE International conference on data mining*. IEEE, 2010, pp. 995–1000.

[22] I. Cohen, Y. Huang, J. Chen, J. Benesty, J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," *Noise reduction in speech processing*, pp. 1–4, 2009.

[23] H. Guo, R. Tang, Y. Ye, Z. Li, and X. He, "Deepfm: a factorization-machine based neural network for ctr prediction," *arXiv preprint arXiv:1703.04247*, 2017.

[24] D. García-García, E. P. Hernández, and F. Díaz-de María, "A new distance measure for model-based sequence clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 31, no. 7, pp. 1325–1331, 2008.

[25] Y. Lin and A. Abur, "A highly efficient bad data identification approach for very large scale power systems," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 5979–5989, 2018.

[26] E. Acuna and C. Rodriguez, "The treatment of missing values and its effect on classifier accuracy," in *Classification, Clustering, and Data Mining Applications: Proceedings of the Meeting of the International Federation of Classification Societies*. Springer, 2004, pp. 639–647.

[27] B. Craven and S. M. Islam, "Ordinary least-squares regression," *The SAGE dictionary of quantitative management research*, pp. 224–228, 2011.

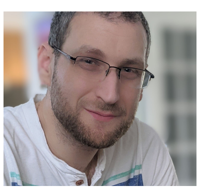
[28] X. He and T.-S. Chua, "Neural factorization machines for sparse predictive analytics," in *40th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2017, pp. 355–364.

[29] X. Wang, F. Yu, and W. Pedrycz, "An area-based shape distance measure of time series," *Applied Soft Computing*, vol. 48, pp. 650–659, 2016.

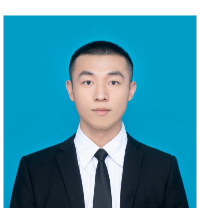
[30] D. García-García, E. Parrado Hernández, and F. Díaz-de María, "A new distance measure for model-based sequence clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 7, pp. 1325–1331, 2009.

[31] M.-F. Guo, W.-L. Liu, J.-H. Gao, and D.-Y. Chen, "A data-enhanced high impedance fault detection method under imbalanced sample scenarios in distribution networks," *IEEE Transactions on Industry Applications*, vol. 59, no. 4, pp. 4720–4733, 2023.

Zhenyu Zhao (S'24) received his B.Eng in Automation from Wuhan University of Technology, China, M.S. in Electrical Engineering from George Washington University, Washington, DC, and Ph.D. in Electrical and Computer Engineering from Temple University, Philadelphia, PA, in 2018, 2020, and 2024, respectively. He is a Postdoctoral Research Associate at the University of Birmingham, UK. He was an intern with Siemens, Minnetonka, MN, from June 2024 to August 2024, and with PJM Interconnection, Audubon, PA, from June 2023 to April 2024. His research interests include advanced data analytics and machine learning methods with applications to transmission systems.



Daniel Moscovitz has 20+ years' experience with PJM Interconnection with a focus on transmission network applications and energy management system (EMS) support. He has earned a Bachelor of Science in Electrical Engineering from Bucknell University, Lewisburg, PA, a Master of of Science in Electrical from Drexel University, Philadelphia, PA, and is currently a Ph.D. candidate at Temple University, Philadelphia, PA. His research focuses on bulk electric grid optimization in the renewable energy supply driven future through improvements in behind the meter solar detection and transmission constrained economic dispatch of energy and reserves.



Shengyi Wang (S'17-M'22) received the Ph.D. degree in electrical engineering from Temple University, Philadelphia, PA in 2022. He is currently a Postdoctoral Fellow Research Associate and Lecturer at Temple University. He received two best paper awards from IEEE PES General Meeting. His research focuses on AI-centric, data-driven solutions for estimation, control, and aggregation in power systems with large-scale distributed energy resources.

Liang Du (S'09–M'13–SM'18) received the Ph.D. degree in electrical engineering from Georgia Institute of Technology, Atlanta, GA in 2013. He was a Research Intern at Eaton Corp. Innovation Center, Mitsubishi Electric Research Labs, and Philips Research N.A. in 2011, 2012, and 2013, respectively. He was an Electrical Engineer with Schlumberger, Sugar Land, TX, from 2013 to 2017. He is currently an Associate Professor of Electrical and Computer Engineering at Temple University, USA. Dr. Du received ORAU's Ralph E. Powe Junior Faculty Enhancement Award in 2018, National Academy of Science Early-Career Fellowship in 2022, and National Science Foundation CAREER award in 2023. He currently serve as an associate editor for IEEE TRANSACTIONS ON SUSTAINABLE ENERGY and IEEE TRANSACTIONS ON TRANSPORTATION ELECTRIFICATION.

Xiaoyuan Fan (S'12–M'16-SM'19) received the Ph.D. degree in electrical engineering from the University of Wyoming, Laramie, WY, in 2016, and M.S. and B.S. degrees in electrical engineering from Huazhong University of Sciences & Technology, Wuhan, China, in 2012 and 2009, respectively. He is currently a Senior Staff Engineer and Power Electronics Team Leader with the Pacific Northwest National Laboratory (PNNL), Richland, WA, USA. His research interests focus on data analytics for power system reliability, multi-discipline resilience analysis and high-performance computing. He is a Senior Member of IEEE, and serves as a volunteer reviewer of 20+ top-level journals and conferences in the area of power systems and signal processing.