A Pivot-point-based Time-series Feeder Load Disaggregation Method

Jiyu Wang, *Member, IEEE*, Xiangqi Zhu, *Member, IEEE*, Ming Liang, *Member, IEEE*, Yao Meng, *Member, IEEE*, Andrew Kling, David Lubkeman, *Fellow, IEEE*, and Ning Lu, *Senior Member, IEEE*

[1]

***Abstract*—A feeder-head load profile is usually known to utility engineers but nodal load profiles on the feeder are not. Thus, in this paper, we present a pivot-point based, two-stage feeder load disaggregation algorithm using smart meter data. The two stages are load profile selection (LPS) and load profile allocation (LPA). In the LPS stage, a few pairs of pivot points are first selected as the matching targets. Then, a random load profile selection process is executed to meet the load diversity requirement. After that, a matching algorithm will run repetitively and select one load profile at a time to match the reference load profile at the pivot points while satisfying load composition requirements. In the LPA stage, the LPS selected load profiles are allocated to each load node on the feeder considering distribution transformer loading limits, load composition, and square-footage. The proposed method is developed and validated using actual feeder and smart meter data from a North Carolina service area. Three performance criteria are used to measure the performance of the algorithm: accuracy of selecting the total number of load profiles, accuracy of matching load shapes, and accuracy of meeting load composition and total nodal square footage requirements. Simulation results show that the proposed method provides more accurate feeder simulation results than the existing load profile allocation methods.**

***Index Terms*— distribution system, feeder load disaggregation, quasi-static time-series analysis, smart meter data**

NOMENCLATURE

A_i	Average square footage of buildings supplied by node i
$A_{i,k}$	Square footage of building k supplied by node i
A_m	Square footage of the building to be allocated
$\Delta A_{i,m}$	Square footage differential percentage between node i and the building to be allocated
D	Load profile database
D_{LT} (D_R, D_C, D_I)	Load profile database for a load type (residential, commercial and industrial)
D_{LSI}	Load profile set containing all load profiles with the largest similarity index
$D_{selected}$	Load profiles selected by LPS
e	Aggregated percentage error between the actual feeder-head load profile and the profile generated by simulation
f	Flag to check if the load profile can be allocated to the feeder
i	Index of the load node
j	Index of the pivot-point pair
J	Total number of pivot-point pairs
k	Index of the selected load profiles at node i
m	Index of load profiles
N_D	Number of load profiles in load pool D
N_i	Number of load profiles at load node i
N_i^{min}, N_i^{max}	Maximum/minimum customers at node i
N_L	Number of load nodes in the feeder
N_{match}	Number of load profiles selected in the shape matching step
N_{random}	Number of load profiles selected in the random selection step
N_T	Matching time duration
N_{feeder}	Total number of selected profiles for the feeder
$N_{selected}$	Number of load profiles selected by LPS
n_{peak}^j, n_{valley}^j	Indices of peak/valley pivot points in X_m
P_{feeder}	Actual feeder-head load profile
$\tilde{P}_{feeder}$	Generated feeder-head load profile
P_D	Aggregated load profile of D
P_{random}	Aggregated profile for the load profiles selected in the random selection process
P_{random}^{target}	Target peak of P_{random}
P_{match}	Aggregated load profile of load profiles selected in the shape matching process
P_{tar}	Target load profile for shape matching
P_{ref}	Reference load profile
$\tilde{P}_{LT}$ ($\tilde{P}_R$,$\tilde{P}_C$ $\tilde{P}_I$)	Aggregated load profile of the selected load profiles that belong to a certain load type (residential, commercial or industrial)
P_m	The m^{th} load profile
$\tilde{P}_i$	Aggregated load profile at load node i
$P_{i,k}$	Load profile k at load node i
q	Index for the first procedure in LPA
φ_m	Similarity index of house m
R_i	Maximum loading factor of transformer i
R_{LT} (R_R,R_C,R_I)	Target ratio of a certain load type (residential, commercial or industrial)
R_{random}	Randomization ratio
S_i	Distribution transformer rating at node i

This study is funded by the U.S. Department of Energy SunShot Initiative and Center for Advanced Power Engineering Research (CAPER). The project team wants to especially thank Dr. Guohui Yuan, Dr. Kemal Celik, Ms. Rebecca Hott, and Mr. Jeremiah Miller from the Systems Integration Subprogram at the Department of Energy SunShot Initiative for their continuing support, help, and guidance. The authors would also like to thank Steve Whisenant and John Gajda at Duke Energy for providing the team with the data and feedback.

J. Wang, X. Zhu, M. Liang, Y. Meng, Dr. Lubkeman and Dr. N. Lu are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606. (e-mails: {jwang49, xzhu9, mliang3, ymeng3, dllubkem, nlu2}@ncsu.edu). Andrew Kling is with Duke Energy Corporation, Charlotte, NC 28202. (e-mail: Andrew.Kling@duke-energy.com.)

$s(m)$	Decision variable representing whether load profile m is selected or not.
t	Index of time step
$t_{peak}^{pp,j}$	j^{th} pivot-point that requires more load to compensate at this time point
$t_{valley}^{pp,j}$	j^{th} pivot-point that requires less load to compensate at this time point
X_m	Matrix for similarity index calculation
u	Number of iterations in RT-LPM/ RNR-LPM
ε	Acceptable error tolerance

I. Introduction

QUASI-STATIC Time-Series (QSTS) power flow simulations have been widely used for feeder modeling in power distribution system planning studies [1-4] in recent years. The most critical step in a QSTS simulation is the selection of nodal load profiles. In earlier QSTS studies, researchers use the shape of the feeder-head load profile as a template to populate all nodal load on the same feeder by normalizing the load profile by the peak load followed and then scaling up to nodal load profiles by corresponding distribution transformer ratings [5-8]. As a result, all load nodes on the feeder have the same load shape. This one-load-shape-scaling (OLSS) method has two main drawbacks. First, because all load nodes have the same load shape, there is no diversity in nodal load changes, i.e., all nodal loads will increase and decrease simultaneously, the simulation results are not realistic. Second, because the nodal load profile does not reflect any actual end use characteristics, it is impossible to model DERs connected to each node.

To improve the OLSS method, Mather [9] and Zhu et al. [10] proposed a multiple-load-shape-scaling (MLSS) method for allocating typical load profiles in different load classes to each load node. For example, street lighting loads are allocated to a small load node and industrial or commercial loads are allocated to nodes with higher distribution transformer ratings. The load profile of each load class is derived from a group of actual load profiles. Thus, MLSS can allocate several classes of typical load profiles. Although there is more diversity in nodal load shapes, MLSS cannot solve the two fundamental drawbacks of OLSS. This is because a typical distribution feeder supplies hundreds or thousands of customers and nodal load profiles can vary drastically when supplying different numbers of customers. Therefore, it is infeasible to use only a limited number of typical load profiles. Another disadvantage of both OLSS and MLSS method is that they only modeled the aggregated profile at each load node instead of providing load profiles for each building located at that node. In DER studies, a modeler needs to know the load profile of each individual building so that impacts of using behind-the-meter loads and DERs (e.g., roof-top photovoltaics (PV) and energy storage devices, electric vehicles, controllable loads, etc.) for providing grid services can be modeled [11,12].

In [13-14], we proposed a load disaggregation method using a load profile database containing hundreds of load profiles. Although this method generates unique nodal load profiles, it only considers using the feeder-head peak load as the pivot point and using distribution transformer ratings as constraints for allocating load profiles. The method fits well for matching daily load profiles without considering load composition, nodal loading ratio, square footage.

In this paper, we described a pivot-point based, two-stage feeder load disaggregation algorithm (FLDA) using smart meter data. The two stages are a load profile selection (LPS) stage followed by a load profile allocation (LPA) stage. In the LPS stage, pivot points are selected from the feeder-head load profile as matching targets. a random load profile selection process is first conducted for meeting the load diversity requirement. Then, a pivot-point matching algorithm is run repetitively such that at each run, the selected load profile provides the best match to the reference load profile at the pivot points while meeting the load composition requirement. In the LPA stage, an allocation algorithm runs repetitively to assign the LPS selected load profiles to each load node. There are three allocation considerations: transformer loading limits, load composition, and building square-footage. The proposed method is developed and validated using real utility feeders and 4302 sets of yearly smart meter data provided by Duke Energy. The following performance criteria are used to measure the performance of the algorithm: total number of houses selected, accuracy in shape matching, and errors in meeting load composition and total nodal square-footage requirements.

There are three main contributions of the paper. *First*, a dynamic reference profile matching method is developed to improve matching accuracy, the success rate for selecting feasible load profiles, and the computational speed of the LPS process. *Second*, a multi-pair-pivot-point matching algorithm is developed to meet the feeder load disaggregation needs for weekly-, monthly- or yearly feeder-head load profiles. *Third*, feeder/nodal load compositions and square-footage requirements are accounted for in the load profile selection and allocation processes.

In the rest of the paper, we will introduce the problem formulation in Section II, present the simulation results and performance evaluation in Sections III and IV, and conclude the paper in Section V.

II. Problem Formulation

This section introduces the problem formulation and design considerations of the FLDA.

A. Data Preparation

At the feeder-head, the total load consumption of the feeder, the ratings of distribution transformers, and nodal load types (i.e. residential, commercial and industrial) are usually known to utility engineers. Most utilities have installed thousands or millions of smart meter, so engineers also have access to end use energy consumptions, $P_m(t_n)$, in many service areas, where $P_m(t_n)$ is the load consumption recorded by the m^{th} smart meter at time t_n. Let $D_{N_D \times N_T}$ be the load profile database that consists of N_D load profiles each with N_T data points. Then we have

$$D_{m,n} = P_m(t_n) \quad (1)$$

In this paper, 4302 sets of smart meter data with 30-minute data resolution are used to form D. Thus, if a daily load profile is

used to form D, $N_D = 4302$ and $N_T = 48$. D can also be further divided into residential (D_R), commercial (D_C) and industrial (D_I) load profile databases.

B. Problem Formulation

As shown in Fig. 1, there are two main steps in FLDA: load profiles selection (FLDA-LPS) and load profile allocation (FLDA-LPA). In the first step, FLDA-LPS will select N_{feeder} load profiles from D to obtain an aggregated load, $\tilde{P}_{feeder}$, which is represented as

$$\tilde{P}_{feeder}(t_n) = \sum_{m=1}^{N_{feeder}} P_m(t_n), \quad (2)$$

The objective function of the FLDA-LPS algorithm is to minimize the error between $\tilde{P}_{feeder}$ and P_{feeder} so we have

$$\min \sum_{n=1}^{N_T} (|P_{feeder}(t_n) - \tilde{P}_{feeder}(t_n)|). \quad (3)$$

After the N_{feeder} load profiles are identified by the FLDA-LPS algorithm, the FLDA-LPA algorithm will assign load profiles to each load node according to the load type and the power rating S_i of the distribution transformer at load node i by meeting the following two constraints:

$$\max\left\{\sum_{k=1}^{N_i} P_{i,k}(t_n)\right\} \le S_i * R_i \quad (4)$$

$$N_i^{min} \le N_i \le N_i^{max} \quad (5)$$

In the following sections, the design considerations of the FLDA are discussed in detail.

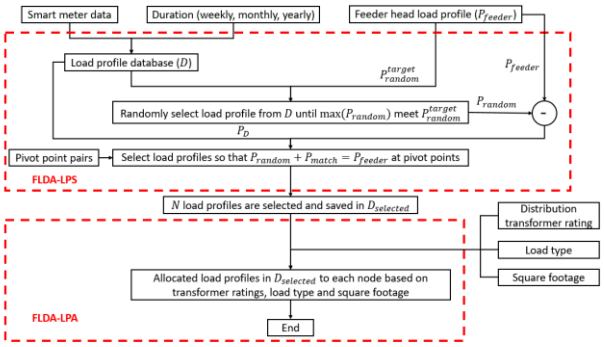

Fig. 1. Flowchart of feeder load disaggregation algorithm

C. FLDA-LPS Algorithm

This section discusses the design considerations of the FLDA-LPS algorithm.

1) Number of Pivot-points

In distribution system studies, it is neither realistic nor necessary to find a perfect match for the entire time series. Therefore, we simplify (3) by minimizing the cumulated error on pivot points:

$$\min \sum_{j=1}^{J} (|P_{feeder}(t^{pp,j}) - \tilde{P}_{feeder}(t^{pp,j})|). \quad (6)$$

As shown in Fig. 2, we developed a pivot-point-based method to capture the load shape information. One pair of pivot points includes a peak and a valley loads and their corresponding timestamps. For matching a daily load profile, selecting one pair of pivot-points is sufficient. However, for longer time-series curves, multiple pairs are needed to provide a wider range of match. Thus, in Section IV, we compare the impacts of selecting different numbers of pivot points on the algorithm performance when disaggregating weekly, monthly and yearly time-series load profiles.

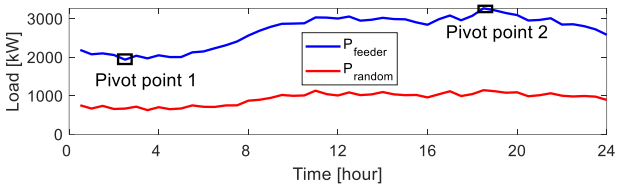

Fig. 2. Pivot points selection and randomly selected loads

2) Random Load Ratio

If all load profiles are selected to satisfy (6) using the pivot-point-based method, all selected load profiles will bear similar load shapes as P_{feeder}. However, on an actual distribution feeder, because of the load diversity, a significant amount of the loads has their unique load shapes. Therefore, we proposed a random load profile selection process to meet the load diversity requirement. Define P_{random} as the aggregated load profile for the load profiles that are random selected and P_{match} as the aggregated load profile for the load profiles that are selected to match $\tilde{P}_{feeder}$ and P_{feeder}, we have

$$\tilde{P}_{feeder}(t_n) = P_{random}(t_n) + P_{match}(t_n) \quad (7)$$

where

$$P_{random}(t_n) = \sum_{m=1}^{N_{random}} P_m(t_n) \quad (8)$$

$$P_{match}(t_n) = \sum_{m=1}^{N_{match}} P_m(t_n) \quad (9)$$

To control the diversity in the load profiles, we introduce a random load ratio, R_{random} to cap the peak of P_{random} by

$$(1-\varepsilon) \times P_{random}^{target} \le \max(P_{random}) \le P_{random}^{target}, \quad (10)$$

$$P_{random}^{target} = R_{random} \times \min(P_{feeder}), \quad (11)$$

where

$$0 \le R_{random} \le 1$$

FLDA randomly selects load profiles from D until (10) is met and save the N_{random} selected load profiles to $D_{selected}$. By capping the maximum load of P_{random} by the minimum load of P_{feeder}, P_{random} will be lower than P_{feeder} when $R_{random} \le 1$. The red curve in Fig. 2 shows an example of P_{random} with $R_{random} = 0.5$. The impact of using different R_{random} values on the algorithm performance is presented in Section III.

3) Recursive Matching Algorithm

After randomly selecting a group of load profiles, FLDA-LPS needs to select the load profiles from D so that $\tilde{P}_{feeder}$ matches P_{feeder}. Thus, a recursive target load profile matching (RT-LPM) method is developed to calculate the target profile for each selection step. The initial target profile is calculated by

$$P_{tar}^{(0)}(t_n) = P_{feeder}(t_n) - P_{random}(t_n) \quad (12)$$

At each selection step, after $P_m^{(u)}$ is selected from D and put into $D_{selected}$, At the end of the u^{th} selection step, $P_{tar}^{(u+1)}$ is recalculated by

$$P_{tar}^{(u+1)}(t_n) = P_{tar}^{(u)}(t_n) - P_m^{(u)}(t_n) \quad (13)$$

At each selection step, based on the calculated $P_{tar}^{(u+1)}$, a pair of pivot points on $P_{tar}^{(u+1)}$ are calculated as

$$P_{tar}^{(u)}(t_{peak}^{pp}) = \max\left(P_{tar}^{(u)}(t_n)\right) \quad (14)$$

$$P_{tar}^{(u)}(t_{valley}^{pp}) = \min\left(P_{tar}^{(u)}(t_n)\right) \quad (15)$$

Based on the pivot points, search D to find a load profile $P_m^{(u)}$ whose valley and peak are at t_{valley}^{pp} and t_{peak}^{pp} or closest to

until (6) is met. The selection criterion is discussed in subsection 4). If multi-pair pivot points are needed, extend (14) and (15) to select multiple peak and valley load pairs.

However, using P_{tar} for selecting load profiles without considering the load characteristics in D has one major drawback. When a load profile P_{match} is selected from D to match the target load profile P_{tar}, statistically, P_{match} with the same or similar load shape as the aggregated load profile of D (See the second subplot of Fig. 3) is more likely to be selected. As a result, $\tilde{P}_{feeder}$ may match P_{feeder} well at points adjacent to the pivot points but poorly at points away from the pivot points.

Therefore, we developed a recursive normalized reference load profile matching (RNR-LPM) algorithm. Normalize P_{tar} by its peak and by the aggregated load profile of D, we obtain a reference profile for step n.

$$P_{ref}^{(u)}(t_n) = \frac{P_{tar}^{(u)}(t_n)/\max(P_{tar}^{(u)}(t_n))}{P_D(t_n)/\max(P_D(t_n))}, \quad (16)$$

where

$$P_D = \sum_{m=1}^{N_D} P_m. \quad (17)$$

As shown in the third subplot of Fig. 3, the pair of pivot points can be calculated as

$$P_{ref}^{(u)}(t_{peak}^{pp}) = \max\left(P_{ref}^{(u)}(t_n)\right) \quad (18)$$

$$P_{ref}^{(u)}(t_{valley}^{pp}) = \min\left(P_{ref}^{(u)}(t_n)\right) \quad (19)$$

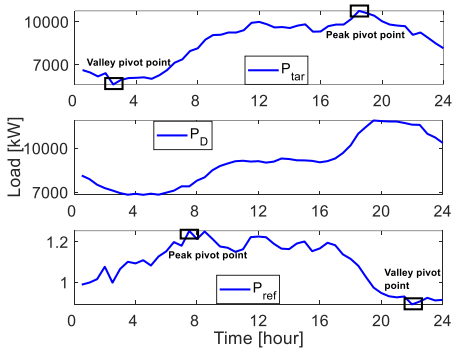

Fig. 3. Pivot points selected using the Target and Reference load profiles.

4) Similarity Index for Load Profile Selection

At each matching selection step, for each load profile in D, form a sorting matrix $\mathbf{X}$ as follows:

$$\mathbf{X_m} = \begin{pmatrix} P_m(t_1) & P_m(t_2) & \dots & P_m(t_{N_T}) \\ t_1 & t_2 & \dots & t_{N_T} \end{pmatrix}. \quad (20)$$

Sort $\mathbf{X_m}$ ascendingly by the first row (i.e., by $P_m(t_n)$), and find n_{peak}^j and n_{valley}^j such that $\mathbf{X_m}\left(2, n_{peak}^j\right) = t_{peak}^{pp,j}$ and $\mathbf{X_m}\left(2, n_{valley}^j\right) = t_{valley}^{pp,j}$, where $t_{peak}^{pp,j}$ and $t_{valley}^{pp,j}$ are the timestamps of the pivot points on the reference curve at that step, as shown in the third subplot of Fig. 3.

Then, calculate the similarity index using

$$\varphi_m = \sum_{j=1}^{J}(n_{peak}^j - n_{valley}^j). \quad (21)$$

Load profile with a larger φ_m has a higher load at $t_{peak}^{pp,j}$ or a lower load at $t_{valley}^{pp,j}$.

As shown in Fig. 4, the pivot points on the reference profile has t_{peak}^{pp} at 7:30 a.m. and t_{valley}^{pp} at 22:00. After sorting two load profiles ascendingly, we found the index of the corresponding load points on the sorted load profile so that φ_1 is calculated as 30 and φ_2 is 46. Thus, Profile 2 has a higher similarity index compared with Profile 1. From the results, we can also see that Profile 2 will provide a better match compared with Profile 1.

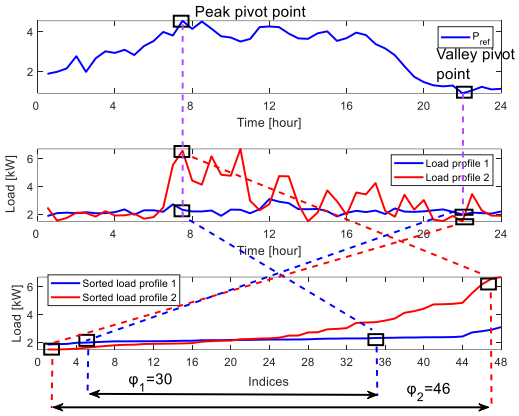

Fig. 4. A graphical illustration of the Similarity Index calculation

5) Tolerance of the Matching Error

Define the matching error tolerance as

$$e = \sum_{t=1}^{N_T}\left|\frac{P_{feeder}(t_n) - \tilde{P}_{feeder}(t_n)}{P_{feeder}(t_n)}\right| \quad (22)$$

In the recursive process, after one load profile is selected, e should decrease. Thus, if a selected load profile causes e to increase, the load profile will be unselected and will be taken out of D.

6) Considerations on Load Types

To consider different load types, D can be divided into different load pools. For example, D_R is the load pool of residential load profiles; D_C is the load pool of commercial load profiles; and D_I is the load pool of industrial load profiles. Instead of directly selecting load profiles from D, load profiles can be selected from one of these three load pools. The selection criteria considering load types are,

$$\frac{\sum_{t=1}^{N_T}\tilde{P}_{LT}(t_n)}{\sum_{t=1}^{N_T}\tilde{P}_{feeder}(t_n)} < R_{LT}, \quad (23)$$

$$\tilde{P}_{LT} = \tilde{P}_{LT} + P_m, \quad (24)$$

$$\frac{\sum_{t=1}^{N_T}\tilde{P}_{LT}(t_n)}{\sum_{t=1}^{N_T}P_{feeder}(t_n)} < R_{LT} * (1+\varepsilon). \quad (25)$$

(23) is used to decide which load type to select. If there are more than 1 load type that satisfies (23), then FLDA will random select one. (24) and (25) ensure that the selected load profiles will not cause the result ratio to exceed the target ratio too much.

7) FLDA-LPS Result Evaluation Criterion

Three criteria are used for evaluating the performance of the FLDA-LPS algorithm: error in matching the feeder-head load shape (ME), error in selecting the total number of load profiles (NE), and error in meeting load composition requirements (LCE). The criteria are calculated by

$$ME = \frac{\sum_{t=1}^{T_D}\left|\frac{P_{feeder}(\mathrm{t}) - \tilde{P}_{feeder}(\mathrm{t})}{P_{feeder}(\mathrm{t})}\right|}{T_D}, \quad (26)$$

$$NE = \left|\frac{N_{feeder} - \tilde{N}_{feeder}}{N_{feeder}}\right|, \quad (27)$$

$$LCE = \left| \frac{\sum_{t=1}^{T_D} \tilde{P}_{LT}(t)}{\sum_{t=1}^{T_D} \tilde{P}_{feeder}(t)} - R_{LT} \right|. \quad (28)$$

When using FLDA-LPS for load disaggregation, one can conduct Monte Carlo simulations to select the best set of results by comparing their evaluation criterion and the robustness of the results, i.e. the deviation (of each MC trial) from their mean statistics. The average results when conducting different number of simulations for a weekly matching are shown in Table I. It shows that at first with the increase of simulation times, the results keep changing, however, the average results become stable once the number of simulations reaches 100. Therefore, in this paper we will run 100 times in Monte Carlo simulation to get the LPS results

Table I. LPS results with different number of simulations

	Number of simulations				
	10	50	100	500	1000
ME (%)	2.34	1.56	1.40	1.41	1.40
NE (%)	6.32	3.24	2.48	2.47	2.50
LCE (%)	0.68	0.59	0.56	0.57	0.57

8) FLDA-LPS Algorithm Summary

The implementation of the FLDA-LPS algorithm is summarized in Algorithm 1.

Algorithm 1 FLDA-LPS

1: Initialize the algorithm. Define the matching duration N_T for P_{feeder}, number of pivot points pairs, energy consumption ratio of each load type (R_R, R_C, R_I) and the load profiles in D (D_R, D_C, D_I).
2: Abandon all load profiles in D with $\max(P_m) \geq \max(S_i R_i)$.
3: **procedure:** Perform Random Load Profile Selection. At each step, add the selected load profile to $D_{selected}$ and remove the selected load profile from D.
4: **Procedure: RNR-LPM**
5: **repeat**
6: Select load pool D_{LT} based on (23).
7: Calculate $P_{ref}^{(u)}$ using (16) and find all pivot-points from $P_{ref}^{(u)}$.
8: Calculate φ_m for all load profiles in D_{LT}, $\Psi = [\varphi_1, \varphi_2 \dots \varphi_{N_D}]$.
9: Save all load profiles with $\varphi_m = \max(\Psi)$ in D_{LSI}.
10: Random draw a load profile, P_m, from D_{LSI}.
11: Apply (24) and (25) to check if the selected load profile P_m can be used. If "yes", go to the next step; if "no', remove the load profile from D_{LT} and go to step 8.
12: $\tilde{P}_{feeder} = \tilde{P}_{feeder} + P_m$.
13: Calculate e using (22).
14: **if** e increases for the last iteration **then**
15: $\tilde{P}_{feeder} = \tilde{P}_{feeder} - P_m$.
16: Remove this load profile from D_{LT} and go back to step 6.
17: **else**
18: Add the load profile to $D_{selected}$, remove it from D, clear D_{LSI}.
19: **end if**
20: **until** D is empty.
21: **end procedure**

In this algorithm, step 2 is to ensure that we will not select a house with very high peak load. By using the distribution transformer rating as a constraint, the selected load profile is guaranteed to have a node that it can be put in.

D. FLDA-LPA Algorithm

After the load profile selection step is finished, the next step is to allocate the selected load profiles to each load node.

1) Allocation based on Distribution Transformer Rating

Using (4) and (5), we can allocate the load profile in $D_{selected}$ to a load node based on the loading factor and the distribution transformer rating at that node. The detailed FLDA-LPA algorithm is illustrated as follows.

Algorithm 2 FLDA-LPA

1: Initialize all load node profiles to be $\tilde{P}_i = [0\ 0\ 0 \dots 0]$ (N_T zeros).
2: Initialize i=1, q=1, all N_i to be 0.
3: **procedure** ALLOCATE ONE LOAD PROFILE TO EACH LOAD NODE
4: **repeat**
5: Draw the load profile $P_{i,k}$ with the q^{th} highest peak load in $D_{selected}$.
6: **Repeat**
7: **if** $\max(\tilde{P}_i + P_{i,k}) > S_i R_i$ **then**
8: $q = q + 1$.
9: **else**
10: $\tilde{P}_i = \tilde{P}_i + P_{i,k}$, $N_i = N_i + 1$, $i = i + 1$, $q = 1$.
11: Remove this load profile from $D_{selected}$.
12: **end if**
13: **until** $\max(\tilde{P}_i + P_{i,k}) < S_i R_i$.
14: **until** $i > N_L$
15: **end procedure**
16: **procedure** ALLOCATE OTHER LOAD PROFILES
17: $i = 1$, $f = 1$.
18: **repeat**
19: Draw the load profile $P_{i,k}$ with the highest peak load in $D_{selected}$.
20: **repeat**
21: **if** $\max(\tilde{P}_i + P_{i,k}) > S_i R_i$ **then**
22: $i = i + 1$, $f = f + 1$.
23: **else**
24: $\tilde{P}_i = \tilde{P}_i + P_{i,k}$, $N_i = N_i + 1$, $i = i + 1$, $f = 1$.
25: Remove this load profile from $D_{selected}$.
26: **end if**
27: **if** $i > N_L$ **then**
28: $i = 1$.
29: **end if**
30: **if** $f > N_L$ **then**
31: Abandon this load profile and go back to find two or more load profiles in D with the aggregated peak matching the peak of the abandoned load profile. Save the newly selected load profile in $D_{selected}$.
32: $f = 1$.
33: **end if**
34: **until** $\max(\tilde{P}_i + P_{i,k}) < S_i R_i$
35: **until** $D_{selected}$ is empty.
36: **end procedure**

In FLDA-LPA, Steps 4-14 guarantee that each load node has at least one load profile; Steps 16-36 allocate the remaining load profiles to each node; f is a flag parameter to record how many nodes the algorithm has tried but with failed allocation. When a load profile cannot be allocated to any load nodes, Steps 30-33 will abandon the load profile and go back to find several smaller load profiles with an aggregated peak matching the abandoned load profile. Because the load shape of one load profile cannot significantly influence the aggregated feeder-head profile, the aggregated feeder-head load profile will still match the actual feeder-head load profile after replacing load profiles that cannot fit into any load nodes.

2) Load Types and Building Square Footage Considerations

If the load type and building square footage are also used for load profile allocation, in step 16-36 of FLDA-LPA, we will only allocate those loads that have the required load type and within the given square footage range to the node. The selection criteria considering square footage are

$$A_i = \frac{\sum_{k=1}^{N_i} A_{i,k}}{N_i}, \quad (29)$$

$$\Delta A_{i,m} = \left|\frac{A_i - A_m}{A_m}\right|, \quad (30)$$

where A_i is the average square footage per building at node i and $\Delta A_{i,m}$ is the allowable deviation from the average values. The load profile with smaller $\Delta A_{i,m}$ will be allocated at node i.

III. FLDA-LPS Performance Evaluation

In this section, performance criteria and simulation results for evaluating the FLDA-LPS algorithm are presented.

A. Simulation Setup

Thirty-minute smart meter data collected from 4302 buildings (3785 residential and 517 commercial) are used so D, D_R, and D_c contains 4302, 3785, and 517 load profiles, respectively. Note that we didn't use the actual feeder load profiles given by the utility for evaluating the algorithm performance. Instead, each time, we randomly select 1000 load profiles from the 4302 buildings based on the given load composition requirement; then, we use the aggregated load profile of the 1000 load profiles as the feeder-head load profile. This is because, by doing so, the actual load profiles in $D_{selected}$ are known so we can compare the FLAD-LPS selected load profiles with the 1000 load profiles to quantify the matching accuracy in how many load profiles are selected and how similar the load profiles are for each load type.

B. Selection of the Ratio of Random Load Profiles

In the FLDA-LPS random selection procedure, R_{random} controls the level of load diversity. To study the influence of the R_{random} value on the FLDA-LPS performance, we increase R_{random} from 10% to 100% at a 10% incremental rate. Running 100 simulations using 100 weekly feeder-head load profiles for each given R_{random}, the obtained statistics of the algorithm performance are summarized in Fig. 5.

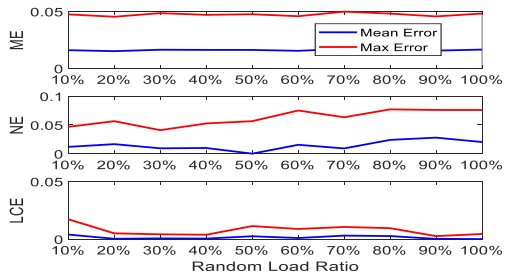

Fig. 5. Performance comparison using different random load ratios.

We observe no obvious performance degradation in ME and LCE when R_{random} changes. This is because R_{random} are selected such that the aggregated randomly selected loads cannot exceed the valley of P_{feeder}. If the peak-to-valley ratio is high for the given feeder-head load profile, the matching algorithm can always find the load profiles from D to make up the error between $\tilde{P}_{random}$ and P_{feeder}. Although there is a little increase in NE because there are larger chances to select extreme cases (more number of smaller or larger load profiles) with the more randomness in selection, the increase is less then 3% from 0 to 100% randomness. Therefore, while applying FLDA-LPS, R_{random} can normally be set to user desired values base on the user's load diversity requirement and the performance of FLDA-LPS will not deteriorate.

C. Selection of Pivot-points

In this section, we will compare the influence of different numbers of pivot-point pairs on FLDA-LPS performance. Set the random load ratios to be 50% and run each case with a given number of pivot-point pairs 100 times.

The simulation results for weekly, monthly, and yearly matching are summarized in Tables II-IV, respectively.

Table II. Statistical result of weekly matching

	Pivot-points pair number						
	1	2	3	4	5	6	7
ME (%)	1.70	1.43	1.42	1.21	1.16	1.64	3.00
NE (%)	4.97	2.54	2.55	1.40	0.31	2.05	3.65
LCE (%)	0.59	0.51	0.66	0.56	0.53	0.69	0.71

Table III. Statistical result of monthly matching

	Pivot-points pair number						
	4	8	12	16	20	24	28
ME (%)	1.83	1.74	1.58	1.54	1.49	1.55	1.61
NE (%)	3.25	3.14	2.85	2.78	2.03	3.12	3.77
LCE (%)	0.57	0.65	0.50	0.54	0.41	0.63	0.56

Table IV. Statistical result of yearly matching

	Pivot-points pair number				
	4	12	24	36	52
ME (%)	3.88	2.64	2.37	2.14	3.61
NE (%)	3.18	3.39	2.36	2.48	3.08
LCE (%)	0.62	0.65	0.68	0.58	0.50

From the results, it can be observed that at the beginning, the matching accuracy will increase with the number of pivot-points pair increase, however, after a certain number is passed, the accuracy starts to decrease. The reason for this result is if the pivot-points pair number is small, the house selected from the load pool will not be similar enough with the target reference profile. Nevertheless, too many pivot-points will have negative effect to the load profile selection because it is impossible that one house has the exactly similar shape with the target reference profile.

While comparing the results, we may also find that with matching period is longer and longer, ME will increase but NE will not change much. On the other hand, LCE will not vary much no matter what the matching period or the number of pivot-points is. Therefore, when using FLDA-LPS, after the target matching period is fixed, we only need to find the best pivot-points pair number that can minimize MAE, and the result from the previous tables can be used as reference.

From the simulation results, our recommendations for selecting the number of pivot points are as follows:

- for weekly load profile matching selecting 4-5 pairs of pivot points (i.e., the peak/valley load of 4/5 days);
- for monthly matching, 16-24 pairs of pivot-points (i.e., the peak/valley loads of 4/5 days in each week);

- for yearly matching, 36 pairs of potential pivot-points (i.e., the peak/valley load of 3 weeks in each month).

D. FLDA-LPS Performance Comparison

We evaluate the performance of the proposed FLDA-LPS for disaggregating a weekly load profile. Method 1 is the FLDA proposed in this paper. Method 2 is the random selection method. Step 2 in Algorithm 1 are used to randomly select the load profiles until e in (22) reaches its minimum value. Method 3 is choosing pivot points with (14) and (15) (RT-LPM) instead of using (18) and (19) (RNR-LPM) for load shape matching procedure. Method 4 is to use a pivot-area method instead of pivot-point for the shape matching procedure. The average energy consumption of three consecutive time steps are calculated and used for the load profile selection. Method 5 is to use an optimization method for load profile selection, which is formulated as,

$$\min \sum_{t=1}^{T_D} \left| P_{feeder}(\mathrm{t}) - \tilde{P}_{feeder}(\mathrm{t}) \right| \quad (31a)$$

$$\text{s.t.: } \tilde{P}_{feeder} = \sum_{m=1}^{N_D} (s(m) * P_m), \quad (31b)$$

$$s(m) = 0, 1\ , \quad (31c)$$

where $s(m)$ is a decision variable representing whether load profile m is selected or not.

In the comparison, for Method 1 and 2 we use the mean value obtained by Monte Carlo simulations. The random load ratio was selected as 50% and 5 pairs of pivot-points were selected. For Method 5, GAMS is used to solve the problem. Because $LTRE$ in FLDA-LPS is consistently small and negligible, it is not selected as a comparison criterion.

As shown in table IV, the performance of FLDA-LPS is the best. Both ME and NE indices of other methods are worse than FLDA-LPS.

Table V. Performance comparison

	Method				
	FLDA	Random	RT-LPM	Pivot-area	GAMS
ME (%)	1.16	4.68	1.98	2.37	2.39
NE (%)	0.31	16.22	7.06	2.54	6.2

E. Performance using different Load Profile Database

This sub-section compares the performance of FLDA-LPS when using different load profile databases. In Case 1, a load profile database from the same area (i.e., 4302 load profiles in North Carolina) as the feeder is used. In Case 2, a load profile database of a different area (i.e., 1500 load profiles in Texas) is used. The parameters are chosen to be same for the two cases, i.e., conducting weekly load profile matching using 5 pairs of pivot-points with R_{rd} at 0.5.

Run each case 100 times. As shown in Fig. 6, as expected, ME is much smaller when using the North Carolina database. This is because the load characteristics are different for even the same type of loads in the two areas. This case shows that the expected errors when using load profiles collected from areas with different load patterns.

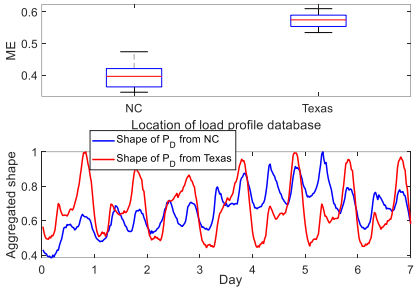

Fig. 6. Performance comparison using two different load profile databases

IV. FLDA-LPA Performance Evaluation

To compare the FLDA-LPA method, we first run FLDA-LPS to select the load profiles by matching the actual, monthly feeder-head load profile of a Duke Energy distribution feeder. Actual ratings of 421 distribution transformers on that feeder are used as an input of FLDA-LPS. Set the peak of the node-level load profile to be less than 80% of the transformer rating. Run FLDA-LPS for 100 times and compare the obtained 100 sets of $D_{selected}$ to select one set of parameters with the minimum ME and $N_{selected}$ closest to the mean. The weighting between the two criterion is 50% for each. Thus, 1295 houses are selected. The results of FLDA-LPA are shown in Figs 6-8.

As shown in Fig. 7, the first subplot shows that the shape of P_{feeder} and $\tilde{P}_{feeder}$ are very close with peaks and valleys closely captured. The second subplot shows the nodal load profiles has significant load diversity. Even for nodes with the same distribution transformer rating, the shapes of the nodal load profiles are different. This will make the distribution system planning study more realistic than using one or a few typical load profiles.

The transformer maximum loading ratio, the number of buildings allocated to, and the load compositions for different transformer size are shown in Fig. 8. The transformer loading levels are within expected range. The number of houses allocated to each node are reasonable. The load compositions are consistent with the expectation of the load composition in the given service area which are verified by Duke Energy customer information data. Note that commercial building loads have a higher chance to be allocated to transformers with higher ratings due to the transformer loading ratio constraints.

Fig. 9 shows the distribution transformer loading ratios distribution for 421 transformers in a typical day. The boxplot shows that the proposed load disaggregation method will let different transformers have non-coincident load peaks and valleys, making the study results more realistic.

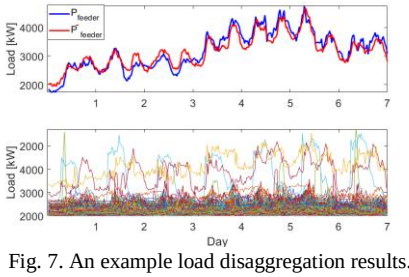

Fig. 7. An example load disaggregation results.

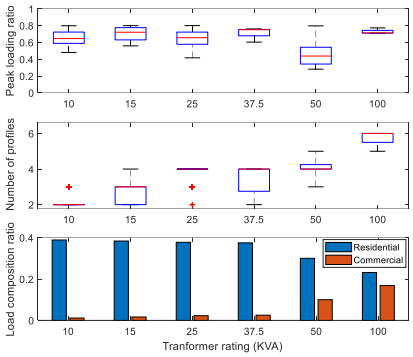

Fig. 8. LPA results for different transformer sizes.

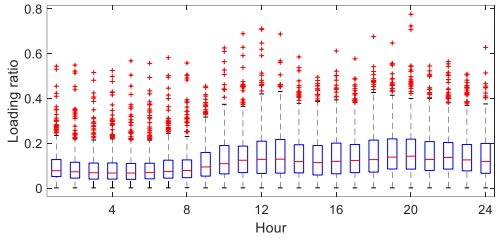

Fig. 9. Transformer loading ratios in an example day.

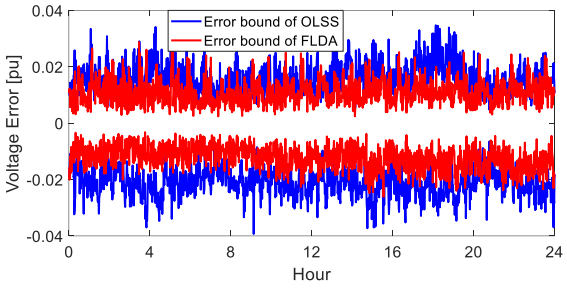

Fig. 10. Nodal voltage error distributions

A comparison between using OLSS and FLDA to conduct time-series power flow analysis is shown in Fig. 10. In the base case, 1000 load profiles are selected and assigned to each node based on the distribution transformer rating at that node. Then, using the aggregated load profile as the feeder top load profile, we use the proposed method to find 1000 load profiles and assign them to each node as the FLDA case. We use the shape of the aggregated load profile to allocated to each node a profile using the OLSS method as the OLSS case.

Power flow studies are conducted for the three cases so that the voltage errors between the OLSS and the base case are compared with the voltage errors between the FLDA and the base case. One set of the results are shown in Fig. 10, where the two blue lines are the upper and lower bounds of the voltage errors for all load nodes using the OLSS approach and the two red lines are the upper and lower bounds of the voltage errors when using the FLDA approach. It can be observed that using the FLDA approach, the voltage profile at each node is closer to the actual compared with using the OLSS approach.

V. Conclusion

The simulation results demonstrated that the pivot-point-based, two-stage FLDA not only ensures that the aggregated load profile of all generated nodal load profiles match the actual metered feeder-head load profile at the pivot points but also provides satisfactory matching results on the rest of the load profile. In addition, modeling information such as the feeder and nodal load composition, the transformer loading ratio, the load diversity ratio, and the nodal supplied square footage requirements can be used as inputs of the load profile selection and allocation processes to give modeler flexibility for modeling different types of operation conditions. By doing so, more realistic QSTS simulations can be conducted on different types of feeders with modeler defined settings. Our future work will be focused on extending the load profile databases to more areas and validate the algorithm performance on more distribution systems.

References

[1] IEEE P1547.7 D5.1, “Draft Guide to Conducting Distribution Impact Studies for Distributed Resource Interconnection,” *IEEE Standard*, Aug. 2011.

[2] M.J. Reno, J. Deboever, and B. Mather, "Motivation and requirements for quasi-static time series (QSTS) for distribution system analysis," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Chicago, IL, USA, 2017 pp. 1-5.

[3] M. Kleinberg, J. Harrison, and N. Mirhosseini, "Using energy storage to mitigate PV impacts on distribution feeders," in *Proc. Power Energy Soc. Innovative Smart Grid Technologies Conference (ISGT)*, Washington, DC, USA, 2014, pp. 1-5

[4] X. Zhu, J. Wang, D. Lubkeman, N. Lu, N. Samaan, and R. Huang, "Voltage-Load Sensitivity Matrix Based Demand Response for Voltage Control in High Solar Penetration Distribution Feeders," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Chicago, IL, USA, 2017 pp. 1-5.

[5] J.A. Jardini, C.M.V. Tahan, S.U. Ahn, and E.L. Ferrari "Distribution transformer loading evaluation based on load profiles measurements," *IEEE Trans. on Power Delivery*, vol.12, no.4, pp. 1766-1770, Oct. 1997.

[6] A. Arif, Z. Wang, J. Wang, B. Mather, H. Bashualdo and D. Zhao, "Load Modeling—A Review," in *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5986-5999, Nov. 2018.

[7] M.J. Reno, and R.J. Broderick. "Predetermined time-step solver for rapid quasi-static time series (QSTS) of distribution systems," in *Proc. IEEE Power Energy Soc. Innovative Smart Grid Technologies Conference (ISGT)*, Washington, DC, USA, 2017, pp. 1-5.

[8] L. Sun, J. Thomas, S. Singh, D. Li, M. Baran, D. Lubkeman, J. DeCarolis, A. Queiroz, L. White, and S. Watts, “Cost-Benefit Assessment Challenges for a Smart Distribution System: A Case Study,” in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Chicago, IL, USA, 2017, pp. 1-5.

[9] B. Mather, "Quasi-static time-series test feeder for PV integration analysis on distribution systems," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, San Diego, CA, USA, 2012, pp. 1-5.

[10] X. Zhu, J. Yan, and N. Lu., "A probabilistic-based PV and energy storage sizing tool for residential loads," in *Proc. IEEE Power Energy Soc. Trans. Distr. Conf.*, Dallas, TX, May. 2016, pp. 1-5.

[11] M.H. Albadi, and E.F. El-Saadany. "Demand response in electricity markets: An overview," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Chicago, IL, USA, 2017 pp. 1-5.

[12] K. Christakou, D.C. Tomozei, J.Y. Le Boudec, and M. Paolone, "GECN: Primary voltage control for active distribution networks via real-time demand-response." *IEEE Trans. Smart Grid*, vol.5, no. 2, pp. 622-631, Mar. 2014.

[13] J. Wang, X. Zhu, D. Lubkeman, N. Lu, N. Samaan, and B. Werts, “Load aggregation methods for quasi-static power flow analysis on high PV penetration feeders,” in *Proc. IEEE Power Energy Soc. Trans. Distr. Conf.*, Denver, CO, Apr. 2018, pp. 1-5.

[14] J. Wang, X. Zhu, D. Mulcahy, C. McEntee, D. Lubkeman, N. Lu, N. Samaan, B. Werts, and A. Kling. "A Two-Step Load Disaggregation Algorithm for Quasi-static Time-series Analysis on Actual Distribution Feeders," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Portland, OR, USA, 2018, pp. 1-5.

[15] PECANSTREET website: https://www.pecanstreet.org/dataport/