Mitigating Smart Meter Asynchrony Error via Multi-Objective Low Rank Matrix Recovery

Yuxuan Yuan, *Graduate Student Member, IEEE*, Kaveh Dehghanpour, and Zhaoyu Wang, *Senior Member, IEEE*

***Abstract*—Smart meters (SMs) are being widely deployed by distribution utilities across the U.S. Despite their benefits in real-time monitoring. SMs suffer from certain data quality issues; specifically, unlike phasor measurement units (PMUs) that use GPS for data synchronization, SMs are not perfectly synchronized. The asynchrony error can degrade the monitoring accuracy in distribution networks. To address this challenge, we propose a principal component pursuit (PCP)-based data recovery strategy. Since asynchrony results in a loss of temporal correlation among SMs, the key idea in our solution is to leverage a PCP-based low rank matrix recovery technique to maximize the temporal correlation between multiple data streams obtained from SMs. Further, our approach has a novel multi-objective structure, which allows utilities to precisely refine and recover all SM-measured variables, including voltage and power measurements, while incorporating their inherent dependencies through power flow equations. We have performed numerical experiments using real SM data to demonstrate the effectiveness of the proposed strategy in mitigating the impact of SM asynchrony on distribution grid monitoring.**

***Index Terms*—Smart meters, sensor asynchrony, low rank matrix recovery, multi-objective optimization.**

Nomenclature

BCSE	Branch current state estimation
DSSE	Distribution system state estimation
MPE	Mean percentage error
PCP	Principle component pursuit
PCA	Principle component analysis
SM	Smart meter
WLS	Weighted least squares
G	Gain matrix
H	Jacobian matrix
h_i	Measurement function that maps state values to the measurement variable i
$\boldsymbol{I_{re}}, \boldsymbol{I_{im}}$	Current real and imaginary values for all the branches
J	Sum of squared residuals
L	Weight parameter for penalizing deviations from SM measurements
M_U	Voltage observation matrix
M_U^*	Refined post-mitigation voltage matrix
M_P	Nodal active power injection matrix
M_P^*	Refined post-mitigation active power matrix
M_Q	Nodal reactive power injection matrix
M_Q^*	Refined post-mitigation reactive power matrix
M_{MV}	Synchronized sensor measurements
M_z	Measurement vector
M_{PS}	Pseudo measurements
$P_i(t_j)$	Measured active power at node i at time t_j
$Q_i(t_j)$	Measured reactive power at node i at time t_j
R	Branch resistance matrix of the system
U_0	Squared voltage magnitude of substation
$U_i(t_j)$	Measured voltage magnitude squared at node i at time t_j
W	Weight matrix
X	Branch reactance matrix of the system
$\boldsymbol{x_s}$	System state vector
Y_M, Y_S	Interim matrices using the latest solution updates
Z_M, Z_S	Interim matrices using the full history of the solution trajectory
α, β	Auxiliary matrices
ΔS_U	Asynchrony voltage error matrix
ΔE_U	Voltage measurement error matrix
ΔS_P	Asynchrony active power error matrix
ΔE_P	Active power measurement error matrix
ΔS_Q	Asynchrony reactive power error matrix
ΔE_Q	Reactive power measurement error matrix
$\delta_U, \delta_P, \delta_Q$	Standard deviations of voltage, active power, and reactive power measurement errors
$\Gamma(\cdot,\cdot)$	Differentiable function for low rank matrices
Γ_T	Total approximate sparsity norm for all the SM datasets
$\|\|\cdot\|\|_*$	Nuclear norm operation
$\|\|\cdot\|\|_1$	1-norm operation
$\|\|\cdot\|\|_F$	Frobenius norm operation
$<\cdot,\cdot>$	Frobenius inner product
$\lambda_U, \lambda_P, \lambda_Q$	Balanced parameters for voltage, active power, and reactive power measurements
μ_U, ν_U	Smoothness parameter
$\omega_1, \omega_2, \omega_3$	Non-negative weights
$\Psi(\cdot,\cdot)$	Differentiable function for sparse error matrices

Manuscript received February 6, 2021; revised April 24, 2021; accepted June 9, 2021. This work was supported in part by the National Science Foundation under Grant EPCN 2042314, and in part by the Advanced Grid Modeling Program at the U.S. Department of Energy Office of Electricity under Grant DE-OE0000875. Paper no. TSG-00198-2021. *(Corresponding author: Zhaoyu Wang.)*

AQ1

The authors are with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011 USA (e-mail: kavehdeh1@gmail.com; wzy@iastate.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2021.3088835.

Digital Object Identifier 10.1109/TSG.2021.3088835

σ_i^2	Error variance of sensor i
$\tau(\cdot,\cdot)$	Aggregate gradient factor
$\zeta_j(A)$	j'th singular value of an arbitrary matrix A.

I. Introduction

The wide-scale deployment of smart meters (SMs) provides a unique opportunity for utilities to enhance their situational awareness capabilities in distribution grids. By 2018, more than 150 million customers across the U.S. were equipped with SMs [1]. On the other hand, SMs are commonly counted among low-quality sensors. Specifically, SMs are asynchronous due to mismatching in sampling time among sensors in the grid, which can limit their applicability in real-time system monitoring [2].

Most previous works on distribution grid state estimation have assumed that SMs are perfectly time-aligned [3], [4]. Only few works have studied the impact of time misalignment and asynchrony of various sensors on grid monitoring and situational awareness: In [5], [6], the statistical characteristics of time misalignment in distribution grid sensors have been estimated using Markov-modulated models. In [2], exponential load variation trends are exploited for developing confidence intervals for SM data samples in distribution system state estimation (DSSE) to compensate for time delays and asynchrony. In [7], a dynamic DSSE formulation is proposed for multitude of asynchronous sensors, which has proven bounded estimation errors. In [8], [9], meter clock synchronization errors are captured through Gaussian probability distributions and represented in DSSE. This idea was also applied in [10] to model measurement errors in grid monitoring. Most solutions proposed for mitigating SM data quality issues rely on *a priori* knowledge of error distribution structure and parameters, which can be difficult to acquire due to information scarcity.

In this paper, we propose a SM data recovery technique that is capable of mitigating the impact of asynchrony error in grid monitoring. Our method has three novel features: (1) We have noted that a rise in SM asynchrony results in a loss of mutual temporal correlation in their time-series data streams. This loss of temporal correlation can be translated into an increase in the rank of *observation matrices*, which store the measurement data from multiple SMs. Thus, we propose to cast the asynchrony error mitigation problem as a low rank matrix recovery process. For this purpose, we have leveraged principle component pursuit (PCP) techniques [11], [12]. PCP employs data-centric optimization for decomposing SM datasets to identify and separate asynchrony error term from raw data. The main idea is that by manipulating the SM data and reducing the rank of the observation matrices, we will enhance the temporal correlation among the SMs which rolls back the adverse impact of asynchrony. (2) In addition to asynchronous errors, SM data has measurement errors that result from the imprecision (i.e., noise) of the measuring devices. Typically, SMs have a relative measurement error of about 1%. Further, unlike image datasets, synchronous SM measurements and asynchronous errors cannot be exactly low rank and exactly sparse. These data properties hinder the applications of state-of-the-art low rank data recovery methods to deal with SM asynchrony errors, such as robust principal component analysis (PCA) [13]. To deal with these problems, we utilize a relaxation to PCP that introduces an entry-wise noise term to represent SM measurement errors in the objective function and eliminate rank-1 constraints. (3) SMs are multi-modal, meaning that they can measure several different variables, including nodal voltage magnitude and nodal average active power (plus nodal reactive power, in some cases.) To mitigate the impact of sensor asynchrony, data recovery needs to be conducted over all measurement datasets simultaneously. However, since these multi-modal datasets are inherently interdependent due to the grid physics, a coordination scheme is required to revise all the datasets while capturing their dependencies. To achieve this, we propose a new multi-objective data recovery formulation that refines voltage magnitude, active/reactive power measurements (and pseudo-measurements), concurrently. The dependencies among these datasets are captured via approximate DistFlow-based constraints [14], [15]. We have developed a Nesterov-based technique to solve the PCP-based multi-objective optimization for recovering multiple SM datasets [16].

The main contributions of this paper are summarized as follows.

- An important observation from real data is presented: asynchrony results in loss of temporal correlation among neighboring SMs. This observation can be quantified using the rank of the nodal voltage observation matrix.
- A novel low rank-based data recovery method is developed to fully mitigate asynchronization error in grid monitoring based on our observation.
- The proposed method considers various specific properties of SM data for enhancing the quality of the recovered data and ensure consistency with grid physics: 1) SMs can measure several different asynchronous variables; 2) SM measurements are statistically interdependent; 3) small entry-wise measurement errors exist within SM measurements.
- Our method handles SM asynchrony issue without needing high-resolution reference sensors, such as micro-PMUs, which are unavailable in most practical distribution systems.
- The proposed solution has been tested using real SM data and feeder models to verify its performance.

The rest of the paper is constructed as follows: Section II presents the proposed multi-objective data recovery method and our approximate first-order solution; Section III demonstrates the application of data recovery in grid monitoring; Section IV analyzes numerical results and verification of the proposed models; finally, Section V presents the paper conclusions.

II. Multi-Objective SM Data Recovery Strategy

In this section, we lay out our data recovery solution for mitigating the errors caused by the asynchronous nature of SMs in distribution grids. This includes key ideas in developing a multi-objective optimization formulation, along with an approximate first-order algorithm to solve the model.

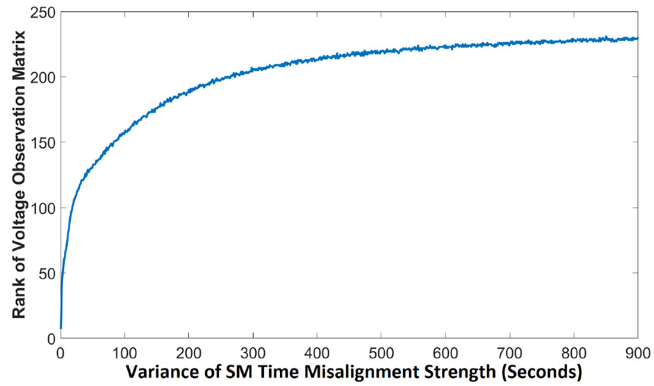


Fig. 1. Rank increase in M_U due to SM asynchrony.

A. Rationale

The available data from SMs can be organized into *observation matrices*. These matrices capture the time-series measurements of several sensors within a given time window $[t_1, t_m]$. For example, the voltage observation matrix is as follows:

$$M_U = \begin{bmatrix} U_1(t_1) & \cdots & U_N(t_1) \\ \vdots & \ddots & \vdots \\ U_1(t_m) & \cdots & U_N(t_m) \end{bmatrix} \quad (1)$$

where, $U_i(t_j)$ is the measured voltage magnitude squared at node i and at time t_j. Note that each column of M_U corresponds to an SM. The observation matrices can be constructed at feeder-, lateral-, or service transformer-levels.

Our PCP-based data recovery model is based on a key observation from real data: asynchrony among SMs leads to an increase in the rank of M_U. The increase in rank is caused by loss of temporal correlation among SMs, which translates into a decrease in statistical correlations in columns of M_U (i.e., the columns lose linear dependency.) This observation can be backed-up by numerical experiments, as shown in Fig. 1. This figure shows the average rank of M_U at various time windows (measured for a grid lateral) as a function of strength of SM asynchrony (measured in terms of variance of time misalignment distribution.) As is observed, the rank of the observation matrix increases as the SM asynchrony intensifies. Note that this observation can be found on the data from SMs with diverse resolutions, including 15 minutes, 30 minutes, and 60 minutes.

To fully capture and mitigate the impact of SM asynchrony, similar observation matrices can be defined for nodal active and reactive power injection measurements, denoted as M_P and M_Q, respectively:

$$M_P = \begin{bmatrix} P_1(t_1) & \cdots & P_N(t_1) \\ \vdots & \ddots & \vdots \\ P_1(t_m) & \cdots & P_N(t_m) \end{bmatrix} \quad (2)$$

$$M_Q = \begin{bmatrix} Q_1(t_1) & \cdots & Q_N(t_1) \\ \vdots & \ddots & \vdots \\ Q_1(t_m) & \cdots & Q_N(t_m) \end{bmatrix} \quad (3)$$

where, $P_i(t_j)$ and $Q_i(t_j)$ are active and reactive power measurements at node i and time t_j, respectively. Note that in general SMs are capable of measuring both average active and reactive powers. However, in many cases, this function is not activated for residential sensors. Thus, in case the reactive power data is unavailable, *pseudo-measurements* can be applied instead to construct an approximate M_Q. Note that our method is robust to gross sparse errors, thus, it can handle the uncertainty of pseudo-measurements and low quality data.

B. Data Recovery Model

The main component of asynchrony error mitigation is to compensate for the loss of temporal correlation among SMs. Since this loss can be detected via the changes in the ranks of the observation matrices, asynchrony error mitigation can be written as a low rank matrix recovery model. To consider both asynchrony errors and small entry-wise measurement errors in SM data, our data recovery approach models an observation matrix (i.e., asynchrony voltage magnitude matrix) as the summation of three components: a low rank voltage magnitude matrix, an asynchrony error matrix, and a measurement error matrix. The goal is to identify unknown voltage magnitude matrix and asynchrony error matrix within the datasets in the presence of entry-wise noise. The model is shown below:

$$M_U = M_U^* + \Delta S_U + \Delta E_U \quad (4)$$

where, M_U^* represents the *refined post-mitigation* voltage magnitude matrix which has a low rank, ΔS_U is the asynchrony error matrix, and ΔE_U represents entry-wise measurement errors. It should be noted that measurement error is different from asynchrony error, as mentioned in previous work [9]. The same representation applies to both active and reactive measurements and pseudo-measurements, as follows:

$$M_P = M_P^* + \Delta S_P + \Delta E_P \quad (5)$$

$$M_Q = M_Q^* + \Delta S_Q + \Delta E_Q \quad (6)$$

where, the sub-components are defined similar to (4). The objective of the data recovery process is to revise the SM data in a way that the ranks of observation matrices are minimized (i.e., temporal correlations among SMs are maximized), while the extent of changes made in the original data is kept at a minimum level. This goal can be represented using three objective functions, corresponding to the available datasets, M_U, M_P, and M_Q, as follows:

$$\begin{cases} f_U = \left\|M_U^*\right\|_* + \lambda_U \|\Delta S_U\|_1 \\ f_P = \left\|M_P^*\right\|_* + \lambda_P \|\Delta S_P\|_1 \\ f_Q = \left\|M_Q^*\right\|_* + \lambda_Q \left\|\Delta S_Q\right\|_1 \end{cases} \quad (7)$$

where, $||\cdot||_*$ and $||\cdot||_1$ are the nuclear norm and 1-norm (i.e., sparsity norm) operations, respectively. These norms are calculated as follows [17]:

$$||A||_* = \sum_j \zeta_j(A) \quad (8)$$

$$||A||_1 = \max_j \sum_j |A(i,j)| \quad (9)$$

where, $\zeta_j(A)$ denotes the j'th singular value of an arbitrary matrix A. Further, λ_U, λ_P, and λ_Q are tunable parameters that are leveraged to balance out the two competing components of the objective functions: minimizing the rank of the recovered data versus the amount of changes made in the data during the recovery process. Mathematically, this means that by minimizing f_U, f_P, and f_Q, the data recovery process effectively minimizes the ranks of M_U^*, M_P^*, and M_Q^*. At the same time, the changes made in the data are kept small by penalizing the sparsity norm of matrices ΔS_U, ΔS_P, and ΔS_Q. The three objectives f_U, f_P, and f_Q are evaluated over the datasets that are generated by the same system (e.g., same feeder, lateral, or service transformer). However, these three datasets are not independent from each other due to the power flow constraints. Thus, the re-calibration of these three datasets cannot be performed separately using conventional low rank data recovery methods, such as robust PCA and PCP. To address this problem, we propose a multi-objective PCP-based model that can jointly refine three the SM datasets. The objective function minimizes the ranks of recovered data to realize the best achievable SM re-alignment. Moreover, to incorporate the inherent interdependencies of the three objectives, power flow equations are added as the constraints of the model. The proposed multi-objective optimization is as follows:

$$\min_{M_U^*, M_P^*, M_Q^*} \quad \{f_U, f_P, f_Q\} \tag{10}$$

$$s.t. \quad \left\| M_U - M_U^* - \Delta S_U \right\|_F \le \delta_U \tag{11}$$

$$\left\| M_P - M_P^* - \Delta S_P \right\|_F \le \delta_P \tag{12}$$

$$\left\| M_Q - M_Q^* - \Delta S_Q \right\|_F \le \delta_Q \tag{13}$$

$$M_U^* = M_P^* \cdot R^T + M_Q^* \cdot X^T + 1_{m \times N} U_0 \tag{14}$$

where, $|| \cdot ||_F$ denotes the Frobenius norm of matrix, defined as follows:

$$||A||_F = \sqrt{\sum_i \sum_j A(i,j)} \tag{15}$$

In addition, parameters δ_U, δ_P, and δ_Q are the standard deviations of the measurement/pseudo-measurement errors (obtained using knowledge of sensor tolerance or pseudo-measurement confidence intervals), matrices R and X represent the branch resistance and branch reactance of the network, respectively [18]. U_0 is the primary voltage magnitude squared for the transformer to which the SMs are connected. The rationale behind constraints (11), (12), and (13) is that the refined components (i.e., M_U^*, M_P^*, M_Q^*) are not exactly low rank and the asynchrony error components (i.e., ΔS_U, ΔS_P, ΔS_Q) are not exactly sparse. Such soft constraints allow for slight deviations in the recovered data to compensate for SM measurement errors, which are consistent with our knowledge of measurement device confidence levels. Also, these allow utilities to minimize asynchrony error with noisy practical SM data, which particularly pertains to reactive power data that may be unavailable for residential customers. Constraint (14) is obtained from the linear DistFlow in matrix form [15], which can enforce network physics and capture the inherent dependencies among datasets. The goal of this constraint is to ensure that the recovered SM data is feasible in power engineering context.

Our method follows the line of low rank data recovery techniques that have been commonly used in many areas [11]. Unlike the black box methods that lack interpretability, the proposed model has a solid mathematical foundation to recover a low rank SM data matrix in the presence of gross asynchrony errors. Also, the dependencies among the datasets are basically encoded into the solution through a set of linear equality constraints. Such power flow models can be applied for arbitrary distribution systems. Note that the model is extendable to unbalanced systems in a straightforward way (i.e., full three-phase DistFlow is leveraged). Further, the proposed data recovery model makes no assumptions on system topology or load distribution, which ensures the performance of this model in other distribution systems.

C. Solution Strategy

A major challenge in solving the proposed data recovery model is the existence of power flow constraints (14) that hinders the application of the existing closed-form dual solvers [13]. Another complication is that (10) has three non-smooth objective functions, which makes the problem non-differentiable. To efficiently tackle these challenges, we present a first-order Nesterov-like algorithm to solve the proposed multi-objective data recovery framework [19]. The basic idea of our solution is to approximate the non-smooth objectives with differentiable surrogates. By applying this idea, the following surrogate components can be written for f_U [16]:

$$\left\| M_U^* \right\|_* \approx \Gamma\left(M_U^*, \mu_U\right) = \max_{||\alpha||_2 \le 1} < M_U^*, \alpha > - \frac{\mu_U}{2} ||\alpha||_F^2 \tag{16}$$

$$\|\Delta S_U\|_1 \approx \Psi(\Delta S_U, \nu_U) = \max_{||\beta||_\infty \le 1} < \Delta S_U, \beta > - \frac{\nu_U}{2} ||\beta||_F^2 \tag{17}$$

where, α and β are auxiliary matrices, μ_U and ν_U are smoothness parameters, and $<\cdot, \cdot>$ is the Frobenius inner product [17], calculated as follows:

$$<A, B> = \sum_i \sum_j A(i,j) \cdot B(i,j) \tag{18}$$

Note that the non-differentiable norms are replaced with differentiable functions $\Gamma(\cdot, \cdot)$ and $\Psi(\cdot, \cdot)$ in (16) and (17). The Lipschitz constants for the gradients of $\Gamma(\cdot, \cdot)$ and $\Psi(\cdot, \cdot)$ equal $\frac{1}{\mu_U}$ and $\frac{1}{\nu_U}$, respectively. Similar smooth surrogates are defined and calculated for the objectives f_P and f_Q. By adopting this approximate alternative, the objectives in optimization (10) can be rewritten as a single-objective weighted averaging process by using a scalarization method [20]. Since the relaxed problem is convex, the single-objective formulation is guaranteed to track all the Pareto-optimal solutions, given valid weight assignment to the objectives [21]. The single-objective formulation can be rearranged as follows:

$$\min_{M_U^*, M_P^*, M_Q^*} \quad \Gamma_T\left(M_U^*, M_P^*, M_Q^*\right) + \Psi_T\left(\Delta S_U, \Delta S_P, \Delta S_Q\right) \tag{19}$$

$$s.t. \quad (11) - (14) \tag{20}$$

here, the new objective function consists of two component: (I) Γ_T quantifies the total approximate nuclear norm for all the SM datasets:

$$\Gamma\Big(M_U^*, M_P^*, M_Q^*\Big) = \omega_1\Gamma\big(M_U^*, \mu_U\big) + \omega_2\Gamma\big(M_P^*, \mu_P\big) + \omega_3\Gamma\Big(M_Q^*, \mu_Q\Big) \quad (21)$$

where, ω_1, ω_2, and ω_3 are the non-negative user-defined weights assigned to f_U, f_P, and f_Q, respectively. Assigning a larger weight to the objective function indicates that the function has a higher priority compared to a function with a smaller weight. Further, $\omega_1 + \omega_2 + \omega_3 = 1$ needs to hold to ensure Pareto-optimality. (II) Γ_T is the total approximate sparsity norm for all the SM datasets, as follows:

$$\Psi\big(\Delta S_U, \Delta S_P, \Delta S_Q\big) = \omega_1\lambda_U\Psi(\Delta S_U, \nu_U) + \omega_2\lambda_P\Psi(\Delta S_P, \nu_P) + \omega_3\lambda_Q\Psi\big(\Delta S_Q, \nu_Q\big) \quad (22)$$

This new data recovery formulation (19) is both convex and differentiable. Given the new model, the Nesterov algorithm entails the following steps to solve SM data recovery problem:

Step I - Initialization: $k \leftarrow 0$ (counter initialization); $M_U^*(0) \leftarrow M_U$, $M_P^*(0) \leftarrow M_P$, $M_Q^*(0) \leftarrow M_Q$, $\Delta S_U \leftarrow 0_{m\times N}$, $\Delta S_P \leftarrow 0_{m\times N}$, and $\Delta S_Q \leftarrow 0_{m\times N}$ (solution initialization).

Step II - Component-Wise Gradient Calculation: Obtain the gradients of components (16) and (17) for all the objective functions in the data recovery problem. As shown in [19], these gradients can be computed as follows:

$$\nabla\Gamma\big(M_U^*(k), \mu_U\big) = \alpha^*(\mu_U) \quad (23)$$

$$\nabla\Psi(\Delta S_U(k), \nu_U) = \beta^*(\nu_U) \quad (24)$$

where, α^* and β^* are the optimal solutions of (16) and (17), respectively, obtained for the latest values of M_U^* and ΔS_U at iteration k. Similar gradient values can be obtained for surrogate components of active/reactive power data.

Step III - Aggregate Gradient Computation: Insert the obtained gradients in Step II, to form the overall gradient values for the weighted averaging problem (19):

$$\nabla\Gamma_T\Big(M_U^*, M_P^*, M_Q^*\Big) = \omega_1\alpha^*(\mu_U) + \omega_2\alpha^*(\mu_P) + \omega_3\alpha^*\big(\mu_Q\big) \quad (25)$$

$$\nabla\Psi_T\big(\Delta S_U, \Delta S_P, \Delta S_Q\big) = \omega_1\beta^*(\nu_U) + \omega_2\beta^*(\nu_P) + \omega_3\beta^*\big(\nu_Q\big). \quad (26)$$

Step IV - Interim Variable Updates: This step in the algorithm defines and updates several interim variables. These variables will be leveraged in the data refinement step. The idea is to apply gradient descent using the aggregate gradient components, obtained in Step III, while at the same time penalize deviations from the original measurements. Four interim matrices are defined: Y_M, Y_S, Z_M, and Z_S. While Y_M and Y_S are computed using the latest solution updates, on the other hand, Z_M and Z_S are obtained using the full history of the solution trajectory. Accordingly, the update process for $[Y_M, Y_S]$ is a convex and tractable optimization process, as follows:

$$[Y_M, Y_S] = \underset{M,S}{\arg\min} \Big\{ <\nabla\Gamma_T\Big(M_U^*(k), M_P^*(k), M_Q^*(k)\Big), M> + <\nabla\Psi_T\big(\Delta S_U, \Delta S_P, \Delta S_Q\big), S> + \frac{L}{2}\Big(\|\Delta M\|_F^2 + \|\Delta S\|_F^2\Big)\Big\} \quad (27)$$

$$s.t. \ \ (11)\text{–}(14) \quad (28)$$

where, L is a weight parameter used for penalizing deviations from SM measurements. Here, the deviation from the original data are denoted as ΔM and ΔS (e.g., $\Delta M = M - [(M_U^*(0), M_P^*(0), M_Q^*(0)])$. Similarly, a convex optimization process is defined for updating $[Z_M, Z_S]$, considering full solver trajectory:

$$[Z_M, Z_S] = \underset{M,S}{\arg\min} \Big\{\tau(M,S) + \frac{L}{2}\Big(\|\Delta M\|_F^2 + \|\Delta S\|_F^2\Big)\Big\} \quad (29)$$

$$s.t. \ \ (11) - (14) \quad (30)$$

where, $\tau(M,S)$ is an average aggregate gradient factor with respect to solver history, defined as follows:

$$\tau(M,S) = \sum_{i=0}^{k} <\nabla\Gamma_T\Big(M_U^*(k), M_P^*(k), M_Q^*(k)\Big), M> + <\nabla\Psi_T\big(\Delta S_U, \Delta S_P, \Delta S_Q\big), S>. \quad (31)$$

Step V - Data Refinement: Apply a weighted averaging process using the updated interim variables, from Step IV, to refine the SM data. Based on the suggestion in [16], this weighted update process is written as follows:

$$\begin{bmatrix} M_U^*(k+1) \\ M_P^*(k+1) \\ M_Q^*(k+1) \end{bmatrix} \leftarrow \Big(\frac{k+1}{k+3}\Big)Y_M + \Big(\frac{2}{k+3}\Big)Z_M \quad (32)$$

$$\begin{bmatrix} \Delta S_U(k+1) \\ \Delta S_P(k+1) \\ \Delta S_Q(k+1) \end{bmatrix} \leftarrow \Big(\frac{k+1}{k+3}\Big)Y_S + \Big(\frac{2}{k+3}\Big)Z_S. \quad (33)$$

Step V-Iterate and Terminate: $k \leftarrow k+1$ and go to Step II until the maximum number of iterations is reached. Output the refined SM datasets, M_U^*, M_P^*, and M_Q^*, after algorithm convergence.

III. Enhancing Grid Monitoring Robustness to SM Asynchrony Error

Fig. 2 shows how our proposed data recovery technique can be integrated into grid monitoring systems as a pre-processor. The refined data is continuously fed to a branch current state estimation (BCSE) module to monitor the grid states in real-time, including the real and imaginary parts of currents of all branches [22]. The BCSE method leverages a weighted least squares (WLS)-based solver to minimize the sum of squared residuals (J). This problem can be formulated as an optimization task over the distribution network given the recovered data samples M_U^*, M_p^*, M_Q^* from our multi-objective PCP-based model, as follows:

$$\min_{\boldsymbol{x_s}} \ J = \sum_i W_{i,i}(M_z(i) - h_i(\boldsymbol{x_s}))^2$$

$$s.t. \ \ M_z = \begin{bmatrix} M_{MV} \\ M_U^*(:) \\ M_P^*(:) \\ M_Q^*(:) \\ M_{PS} \end{bmatrix}$$

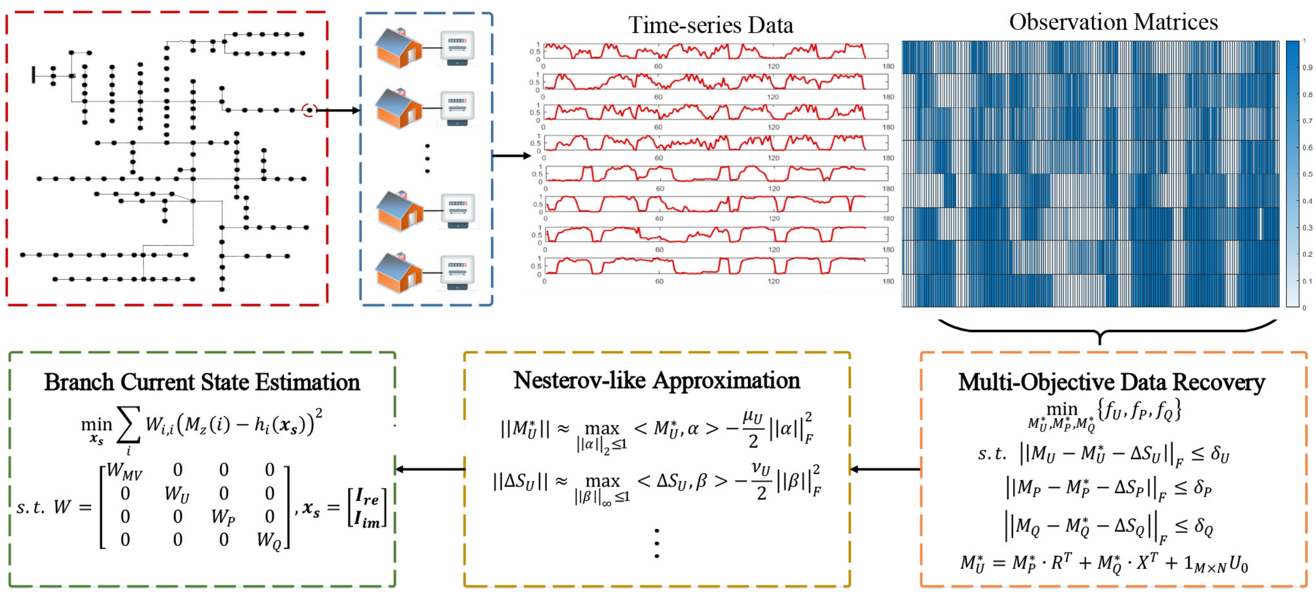


Fig. 2. Overall structure of the solution for grid monitoring.

$$W = \begin{bmatrix} W_{MV} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & W_U & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & W_P & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & W_Q & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & W_{PS} \end{bmatrix}$$

$$\boldsymbol{x_s} = \begin{bmatrix} \boldsymbol{I_{re}} \\ \boldsymbol{I_{im}} \end{bmatrix} \tag{34}$$

where, $\boldsymbol{x_s}$ is the grid state vector that contains current real/imaginary values for all the branches of the distribution system ($\boldsymbol{I_{re}}/\boldsymbol{I_{im}}$), and M_z is the measurement vector. The measurement data includes the MV network synchronized sensor measurements (M_{MV}), including SCADA and μPMUs, if available, the refined SM data, M_U^*, M_P^*, M_Q^*, and the pseudo measurements M_{PS} that can generated by our previous work [23]. h_i is the measurement function that maps state values to the i'th measurement variable, which is obtained based on the power flow equation. Furthermore, W is a weight matrix that represents the solver's confidence level in each element of M_z. The matrix W includes the measurement confidence weights, consisting of sub-matrces W_{MV}, W_U, W_P, W_Q, and W_{PS} corresponding to M_{MV}, M_U^*, M_P^*, M_Q^*, and M_{PS}, respectively. These weight values are determined by the nominal accuracy levels of the senors as $W_{i,i} = \frac{1}{\sigma_i^2}$, where σ_i^2 is the i'th sensor error variance [24]. The purpose of the weights is to devalue the importance of unreliable data sources in grid monitoring.

The WLS-based solution employs a gradient-based algorithm to find the optimal solutions for (34) (i.e., $\nabla_{\boldsymbol{x_s}J=0}$) [25]. The algorithm involves the following steps to estimate the states of the grid:

Step I - Receive Input Data: Receive the recovered SM data, M_U^*, M_P^*, and M_Q^* (see Section II), and the latest measurement data from the primary network, M_{MV}. Concatenate the input data to form the measurement vector, M_z.

Step II - State Initialization: $k \leftarrow 0$; initialize the values of the states through randomization, $\boldsymbol{x_s}(k)$ (to speed up the BCSE solver the values of states can be initialized using the solutions from the last time step.)

Step III - Jacobian Computation: Update the *Jacobian matrix*, H, using the gradients of the measurement function. The Jacobian captures the sensitivity of the measurements to the state variables:

$$H_{i,j} = \frac{\partial h_i(\boldsymbol{x_s}(k))}{\partial \boldsymbol{x_s}_j} \tag{35}$$

The Jacobian matrix can be conveniently calculated for the BCSE method for feeders with known topology (e.g., see [22] for details on how Jacobian can be obtained for various types of measurement functions.)

Step IV - Gain Matrix Computation: Leverage the Jacobian matrix from Step III to obtain the gain matrix, G, as follows:

$$G(\boldsymbol{x_s}(k)) = H^\top(\boldsymbol{x_s}(k))WH(\boldsymbol{x_s}(k)). \tag{36}$$

Step V - State Update: Update the values of the states using the gain matrix within the first order Newton-Gauss method, as follows:

$$\boldsymbol{x_s}(k+1) \leftarrow \boldsymbol{x_s}(k) + G^{-1}H^\top W(M_z - \boldsymbol{h}(\boldsymbol{x_s}(k))). \tag{37}$$

Step VI - Iterate and Terminate: $k \leftarrow k+1$; go back to Step III until convergence, i.e., $k \geq k_{max}$, with k_{max} being a user-defined maximum number of iteration for the BCSE algorithm.

Step VII - Roll the Time Window: At the new time point, the data recovery is performed using the latest measurement data, according to II. Go back to Step I.

IV. Numerical Results

The proposed data recovery and grid monitoring framework has been tested and validated using a fully observable

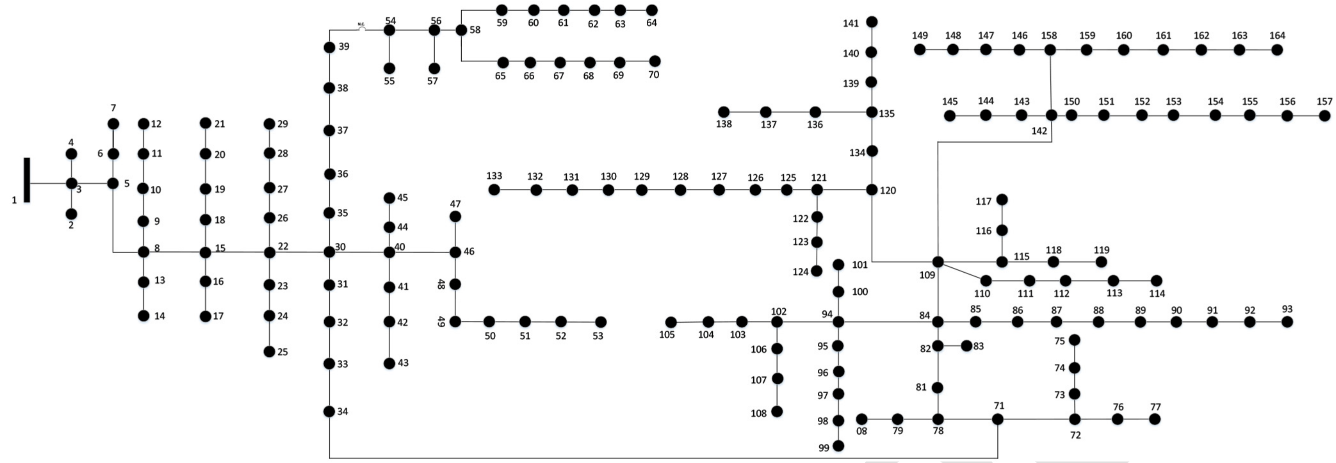


Fig. 3. 164-node feeder topology.

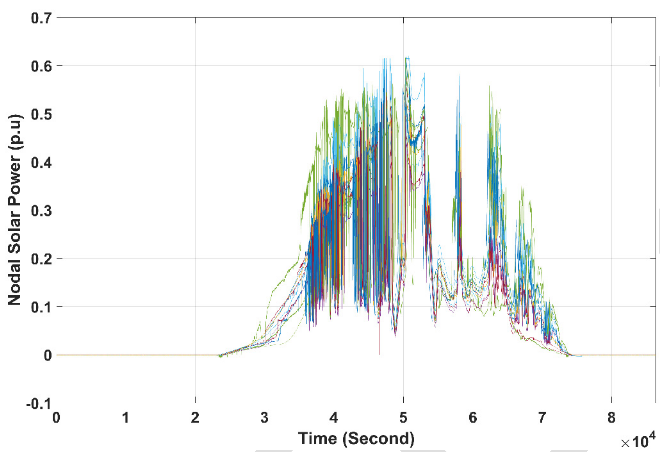


Fig. 4. PV generation data.

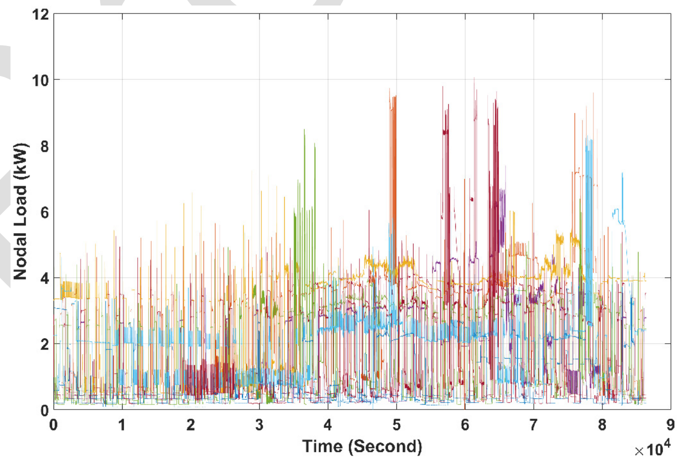


Fig. 5. Consumption data.

feeder model shown in Fig. 3. This feeder represents an unbalanced utility network in U.S. MidWest and consists of 164 nodes, which is publicly available online [26]. The details of the system model include network topology, line parameters, and standard electric components. The system has an average of 30% solar-power-to-peak-load penetration level. The solar data is adopted from [27]. The nodal time-series load demand is aggregated using a real-world 1-second-resolution household dataset and utilized as the input of the power flow analysis [27]. The computed voltages are treated as the voltage measurements. The resolution of the SM measurements is 15-minute. To simulate realistic asynchronous SM measurements, we randomly sample the 1-second resolution data at 15-min rate at each node to represent SM measurements. Thus, in this work, the SM asynchrony strength of each customer can be anywhere between 0 to 900 s. Fig. 4 and 5 show the original solar and load time-series data in a day at different nodes of the system. User-defined parameters within the proposed data recovery model, including coefficients of the optimization solver, have been tuned through try-outs over historical/simulation datasets. Basically, the values of these parameters are chosen when the residual of branch current state estimation is minimized. It should be noted that the high computational budget of this strategy does not impact the real-time performance of the proposed method since this parameter calibration is an offline process.

The case study is conducted on a standard PC with an Intel Xeon CPU running at 3.70 GHz and with 32.0 GB of RAM. Based on 500 Monte Carlo simulations, the average computational time is around 23 s, which is feasible in real-time applications. Fig. 6, 7, and 8 show the average error histograms of the proposed data recovery method for voltage, active power, and reactive power, respectively. The error is calculated by comparing the actual values of various variables with the solutions of the recovery model. As can be observed, the recovery error values are maintained within low levels, which confirms the acceptable performance of the data recovery framework. Specifically, the mean average errors are 0.11%, 2.03%, and 1.27% for voltage, active power, and reactive power, respectively. This also demonstrates that the proposed data refinement framework has the best performance over the SM voltage dataset, among the three datasets. This outcome is consistent with the correlation-driven nature of the data recovery model (i.e., nodal voltage measurements are highly correlated, which facilitates better refinement.)

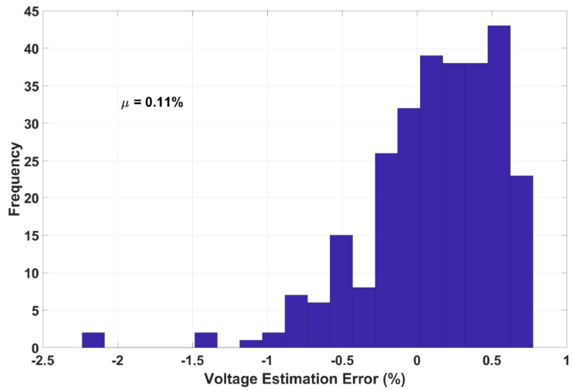


Fig. 6. Voltage recovery error.

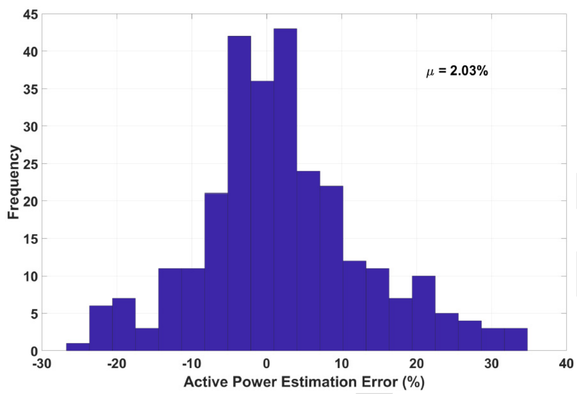


Fig. 7. Active power recovery error.

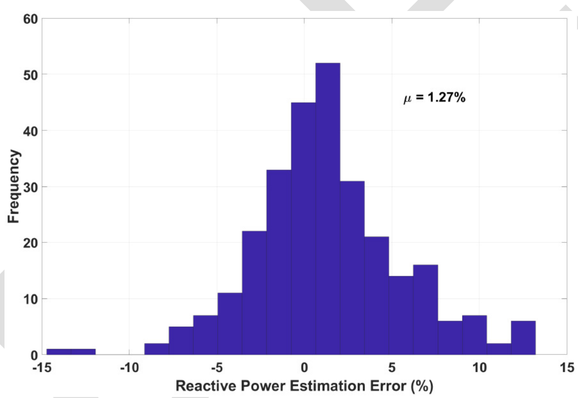


Fig. 8. Reactive power recovery error.

Fig. 9 compares the average value of recovered voltage data from the data refinement framework with the actual nodal voltage average (assuming synchronized sensors) within a sample time-window. As is observed in this figure, the developed algorithm closely follows the underlying signal. Fig. 10 shows a similar concept for active and reactive power datasets. As observed in this figure, the data recovery framework is basically an approximate identity mapping between the recovered data and the underlying (ideal) data. This corroborates the satisfactory performance of the model over real data in time domain.

Fig. 11 shows the histogram of power flow error with and without leveraging the DisFlow equations within the proposed

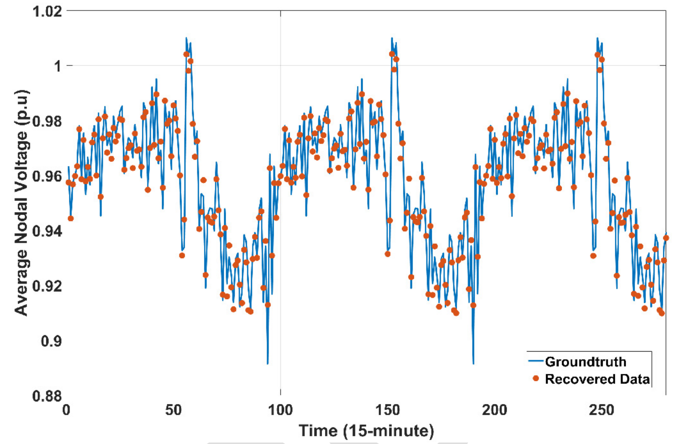


Fig. 9. Recovered average voltage data versus real (synchronized) time-series data.

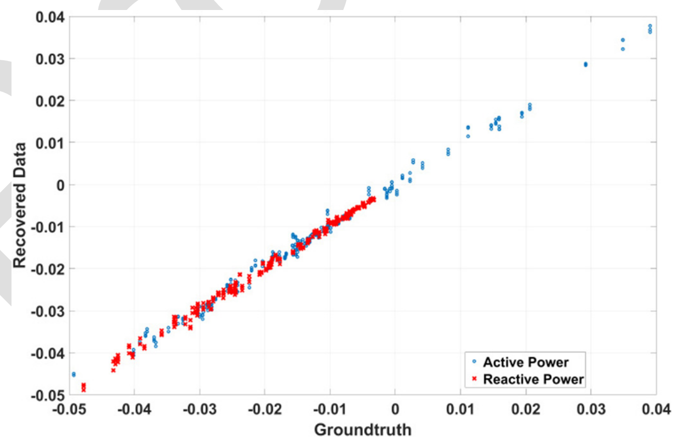


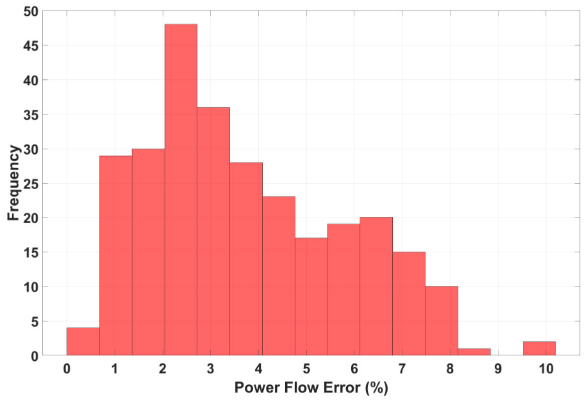
Fig. 10. Recovered nodal active/reactive power data versus real (synchronized) data.

data recovery framework. As can be observed, having the DistFlow equations as constraints within the multi-objective data refinement model has resulted in a significant reduction in power flow errors. This demonstrates that the proposed method is able to output data that is consistent with network physics, while capturing the dependencies among all SM datasets.

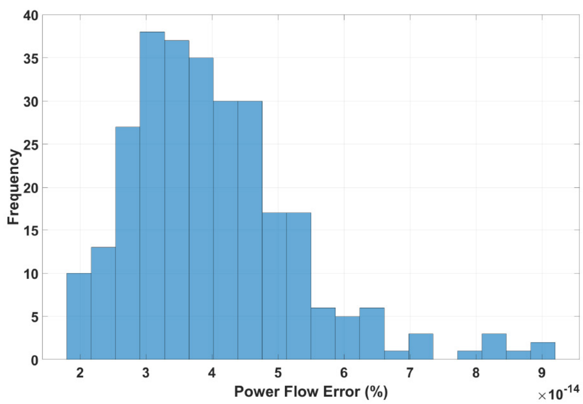
Finally, Fig. 12 depicts the histogram of system monitoring error after applying the data recovery framework. The mean percentage error (MPE) criterion is utilized to evaluate the performance of BCSE with our data recovery method, which is calculated by comparing the real state values ($\boldsymbol{x_s}$), obtained from power flow simulations on the feeder model, with the estimated state values ($\hat{\boldsymbol{x}}_s$), coming from the BCSE, as follows:

$$E = 100 \times \sum_i \frac{\boldsymbol{x_s}(i) - \hat{\boldsymbol{x}}_s(i)}{\boldsymbol{x_s}(i)} \tag{38}$$

As is observed in Fig. 12, the DSSE error value is maintained at low levels, which demonstrates the successful integration of the data recovery solution into grid monitoring, which allows us to track the behavior of the feeder accurately. The mean estimation error value is 0.87%.



(a) Without DistFlow



(b) With DistFlow

Fig. 11. Power flow error with and without DistFlow constraints.

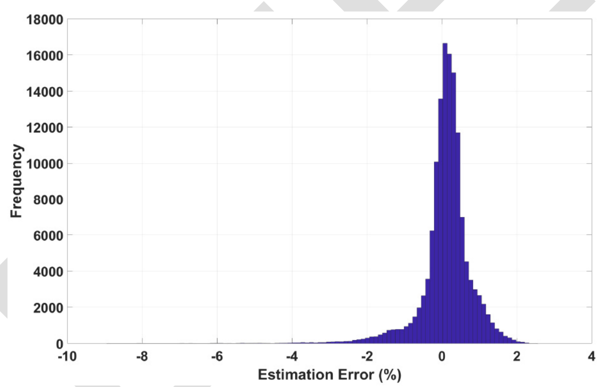


Fig. 12. BCSE error distribution.

To further demonstrate the performance of the proposed SM data recovery method, We have conducted numerical comparisons with two previous methods, including a previous smart meter asynchrony mitigation method [9] and a state-of-the-art low rank data recovery method [11]. The three methods are simulated with the same real-world datasets to calculate the accuracy of the methods. The comparison result is shown in Fig.13. As demonstrated in the figure, in terms of voltage, the average recovery errors are 0.11%, 0.877%, and 1.34% for the proposed solution, [9] and [11], respectively. In terms of active power, the average recovery errors are 2.03%, 5.84%, and 6.48%, respectively. Hence, based on this dataset, the

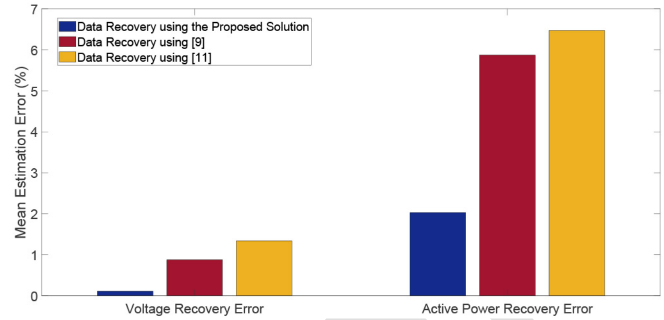


Fig. 13. Comparison results between [9], [11], and the proposed method.

proposed method can achieve a better performance compared to the previous works.

V. Conclusion

In this paper, we have presented a multi-objective data recovery method to mitigate the impacts of SM asynchrony issues in distribution system real-time monitoring. The proposed method is able to refine voltage, active power, and reactive power datasets simultaneously within the same framework via a multi-objective formulation. The inherent dependencies among these measurements are captured by using DistFlow equations. Our solution considers both asynchrony errors and measurement errors, thus making the model more widely applicable to practical distribution systems. A first-order algorithm is presented to solve the proposed multi-objective data recovery model. This algorithm is based on Nesterov method for approximating non-differentiable optimization problems with smooth surrogates. To evaluate the proposed method, a real 164-node utility feeder with real data is utilized. The results show that SM asynchrony error mitigation is possible using the proposed method with good accuracy. In this work, the mean average data recovery error are about 0.11%, 2.03%, and 1.27% for voltage magnitude, active power, and reactive power, respectively. Also, it can be observed that the DistFlow constraints can significantly reduce the inconsistency of recovered data with power flow equations. Based on the proposed data recovery method, the system state estimation error is less than 1%.

References

[1] Energy Information Administration. (2020). *Advanced Metering Count by Technology Type*. [Online]. Available: https://www.eia.gov/electricity/annual/html/epa_10_10.html

[2] J. Lin, P. Wang, S. Guo, Y. Shao, and Y. Sheng, "The credibility modelling and analysis of ami measurements for distribution system state estimation," in *Proc. IEEE Sustain. Power Energy Conf. (iSPEC)*, Beijing, China, 2019, pp. 1556–1560.

[3] A. Primadianto and C.-N. Lu, "A review on distribution system state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 5, pp. 3875–3883, Sep. 2017.

[4] K. Dehghanpour, Z. Wang, J. Wang, Y. Yuan, and F. Bu, "A survey on state estimation techniques and challenges in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 2312–2322, Mar. 2019.

[5] I. Antonios, H.-P. Schwefel, and L. Lipsky, "Approximation of the time alignment error for measurements in electricity grids," in *Proc. 15th Eur. Depend. Comput. Conf. (EDCC)*, Naples, Italy, 2019, pp. 153–158.

[6] H.-P. Schwefel, I. Antonios, and L. Lipsky, “Impact of time interval alignment on data quality in electricity grids,” in *Proc. IEEE Int. Conf. Commun. Control Comput. Technol. Smart Grids (SmartGridComm)*, Aalborg, Denmark, 2018, pp. 1–7.

[7] G. Cavraro, E. Dall’Anese, and A. Bernstein, “Dynamic power network state estimation with asynchronous measurements,” Nat. Renew. Energy Lab. (NREL), Golden, CO, USA, NREL Rep. NREL/CP-5D00-75064, 2019.

[8] S. Bolognani, R. Carli, and M. Todescato, “State estimation in power distribution networks with poorly synchronized measurements,” in *Proc. 53rd IEEE Conf. Decis. Control*, Los Angeles, CA, USA, 2015, pp. 2579–2584.

[9] A. Alimardani, F. Therrien, D. Atanackovic, J. Jatskevich, and E. Vaahedi, “Distribution system state estimation based on nonsynchronized smart meters,” *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2919–2928, Nov. 2015.

[10] F. Ni, P. H. Nguyen, J. F. G. Cobben, H. E. van den Brom, and D. Zhao, “Uncertainty analysis of aggregated smart meter data for state estimation,” in *Proc. IEEE Int. Workshop Appl. Meas. Power Syst. (AMPS)*, Aachen, Germany, 2016, pp. 1–6.

[11] T. Bouwmans and E. H. Zahzah, “Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance,” *Comput. Vis. Image Understand.*, vol. 122, pp. 22–34, May 2014.

[12] P. Rodríguez and B. Wohlberg, “Fast principal component pursuit via alternating minimization,” in *Proc. IEEE Int. Conf. Image Process.*, Melbourne, VIC, Australia, Sep. 2013, pp. 69–73.

[13] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma, “Stable principal component pursuit,” in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, TX, USA, 2010, pp. 1518–1522.

[14] G. M. Gilbert, D. E. Bouchard, and A. Y. Chikhani, “A comparison of load flow analysis using DistFlow, Gauss-Seidel, and optimal load flow algorithms,” in *Proc. IEEE Can. Conf. Elect. Comput. Eng.*, vol. 2. Waterloo, ON, Canada, May 1998, pp. 850–853.

[15] M. E. Baran and F. F. Wu, “Optimal capacitor placement on radial distribution systems,” *IEEE Trans. Power Del.*, vol. 4, no. 1, pp. 725–734, Jan. 1989.

[16] N. S. Aybat, D. Goldfarb, and G. Iyengar, “Fast first-order methods for stable principal component pursuit,” May 2011. [Online]. Available: arXiv:1105.2126.

[17] G. M. Phillips and P. J. Taylor, *Theory and Applications of Numerical Analysis*. Burlington, ON, Canada: Elsevier, 1996.

[18] G. Qu and N. Li, “Optimal distributed feedback voltage control under limited reactive power,” *IEEE Trans. Power Syst.*, vol. 35, no. 1, pp. 315–331, Jan. 2020.

[19] Y. Nesterov, “Smooth minimization of non-smooth functions,” *Math. Program.*, vol. 103, no. 1, pp. 127–152, May 2005.

[20] N. Gunantara, “A review of multi-objective optimization: Methods and its applications,” *Cogent Eng.*, vol. 5, no. 1, 2018, Art. no. 1502242.

[21] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. West Sussex, U.K.: Wiely, 2001.

[22] H. Wang and N. N. Schulz, “A revised branch current-based distribution system state estimation algorithm and meter placement impact,” *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 207–213, Feb. 2004.

[23] Y. Yuan, K. Dehghanpour, F. Bu, and Z. Wang, “A multi-timescale data-driven approach to enhance distribution system observability,” *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 3168–3177, Jul. 2019.

[24] M. E. Baran and A. W. Kelley, “A branch-current-based state estimation method for distribution systems,” *IEEE Trans. Power Syst.*, vol. 10, no. 1, pp. 483–491, Feb. 1995.

[25] R. Singh, B. C. Pal, and R. B. Vinter, “Measurement placement in distribution system state estimation,” *IEEE Trans. Power Syst.*, vol. 24, no. 2, pp. 668–675, May 2009.

[26] F. Bu, Y. Yuan, Z. Wang, K. Dehghanpour, and A. Kimber, “A time-series distribution test system based on real utility data,” in *Proc. North Amer. Power Symp. (NAPS)*, Wichita, KS, USA, 2019, pp. 1–6.

[27] C. Holcomb, “Pecan street inc.: A test-bed for NILM,” in *Proc. Int. Workshop Non-Intrusive Load Monitor.*, 2012, pp. 271–288. [Online]. Available: https://www.pecanstreet.org/

Yuxuan Yuan (Graduate Student Member, IEEE) received the B.S. degree in electrical and computer engineering from Iowa State University, Ames, IA, USA, in 2017, where he is currently pursuing the Ph.D. degree. His research interests include distribution system state estimation, synthetic networks, data analytics, and machine learning.

Kaveh Dehghanpour received the B.Sc. and M.S. degrees in electrical and computer engineering from the University of Tehran in 2011 and 2013, respectively, and the Ph.D. degree in electrical engineering from Montana State University in 2017. He is currently a Postdoctoral Research Associate with Iowa State University. His research interests include application of machine learning and data-driven techniques in power system monitoring and control.

Zhaoyu Wang (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiaotong University, and the M.S. and Ph.D. degrees in electrical and computer engineering from Georgia Institute of Technology. He is the Harpole-Pentair Assistant Professor with Iowa State University. His research interests include optimization and data analytics in power distribution systems and microgrids. He was the recipient of the National Science Foundation CAREER Award, the IEEE Power and Energy Society (PES) Outstanding Young Engineer Award, and the Harpole-Pentair Young Faculty Award Endowment. He is the Principal Investigator for a multitude of projects focused on these topics and funded by the National Science Foundation, the Department of Energy, National Laboratories, PSERC, and Iowa Economic Development Authority. He is the Chair of IEEE PES PSOPE Award Subcommittee, the Co-Vice Chair of PES Distribution System Operation and Planning Subcommittee, and the Vice Chair of PES Task Force on Advances in Natural Disaster Mitigation Methods. He is an Editor of IEEE Transactions on Power Systems, IEEE Transactions on Smart Grid, IEEE Open Access Journal of Power and Energy, IEEE Power Engineering Letters, and *IET Smart Grid*.